A Computer Vision Story on Video Sequences:

From Face Detection to Face Super- Resolution using Face Quality Assessment

Nasrollahi, Kamal

COMPUTER VISION AND MEDIA TECHNOLOGY LABORATORY

PH.D. DISSERTATION

A COMPUTER VISION STORY ON VIDEO SEQUENCES:

FROM FACE DETECTION TO FACE SUPER-RESOLUTION USING FACE QUALITY ASSESSMENT

KAMAL NASROLLAHI

FACULTY OF ENGINEERING AND SCIENCE

AALBORG UNIVERSITY 2011

About the Author

Kamal Nasrollahi received his M.Sc. degree in computer engineering from Amirkabir University of Technology (Tehran Polytechnic), Iran, in 2007. He started his PhD. study in Electrical Engineering at Computer Vision and Media Technology (CVMT) laboratory at Department of Architecture, Design, and Media Technology in Aalborg University, Aalborg, Denmark. He is currently employed as post.doc. at CVMT.

During his master and PhD studies he has been involved in teaching to undergraduate and graduate students. His main research areas include: image processing, computer vision, machine learning algorithms, soft computing algorithms, pattern recognition, and inverse problems. He is a student member of IEEE.

A COMPUTER VISION STORY ON VIDEO SEQUENCES:

FROM FACE DETECTION TO FACE SUPER-RESOLUTION USING FACE QUALITY ASSESSMENT

A PH.D. DISSERTATION BY

KAMAL NASROLLAHI

COMPUTER VISION & MEDIA TECHNOLOGY LABORATORY

FACULTY OF ENGINEERING AND SCIENCE

AALBORG UNIVERSITY, DENMARK

E-MAIL: KN@CREATE.AAU.DK

URL: HTTP://WWW.CVMT.DK/~KN

FEBRUARY 2011

ISBN: 978-87-992732-5-6.

This dissertation was submitted in December 2010 to the Faculty of Engineering and Science, Aalborg University, Denmark, in partial fulfillment of the requirements for the Doctor of Philosophy degree.
The defense took place at Aalborg University, Niels Jernes Vej 14, DK-9220 Aalborg on 09, February 2011. The session was moderated by Associate Professor Hans Jørgen Andersen, Department of Architecture, Design, and Media Technology, Aalborg University.
The following adjudication committee was appointed to evaluate the thesis. Note that the supervisor was a non-voting member of the committee.

Professor Mads Nielsen

Department of Computer Science

University of Copenhagen, Denmark

Associate Professor Patrizio Campisi

Department of Applied Electronics

Università degli Studi, Roma TRE, Italy

Associate Professor Hans Jørgen Andersen (chairman)

Department of Architecture, Design, and Media Technology

Aalborg University, Denmark

Associate Professor Thomas B. Moeslund (Supervisor)

Department of Architecture, Design, and Media Technology

Aalborg University, Denmark

ABSTRACT

Cameras capturing video sequences are nowadays used in almost any public environments like airports, train stations, banks and even in shops. These cameras are usually mounted with wide fields of views in order to cover as much of the scene as possible. Therefore, there are large distances between the cameras and the objects. The immediate consequence is that the objects of interest in such videos are usually very small. One of the most important objects in such video sequences is the human face. However, developing computer vision applications to work with face images in such videos is a challenging task. In the first step, *face detection*, these face images should be detected. Precise face detection for faces of small sizes is itself a difficult task. Though, most of these detected face images are not useable in any computer vision applications. This is due to problems like not facing the camera, blurriness of the face, darkness of the face, small size of the face, and changes in the status of facial components (like closeness of the eyes and openness of the mouth). Using such face images would produce erroneous results in almost any facial analysis system. Furthermore, there are many faces that resemble each other very closely and keeping only one of them may suffice. Therefore, it is necessary to use a mechanism for assessing the quality of the face images, i.e. *face quality assessment*. This mechanism should discard useless face images and summarize the input video sequence to smaller sets containing some of the most expressive images of the video. These summarized sets, containing the most expressive images, are denoted Face-Logs. The decision of the face quality assessment on whether an image in the original video sequence goes into the face-log or not, depends on the application. The proposed system in this thesis is a flexible face quality assessment system which goes through all the required steps for constructing face-logs from video sequences of any length. The system generates face-logs to use them in different computer vision applications. It first constructs a best face-log containing the best face image of the video sequence. This log can be used for video indexing. Then, the system evolves this log by adding *m*-best images of the sequence. Such a face-log can be used as a complete and concise representation of the video sequence for video summarization as well as for recognition purposes. Finally, the system goes one step further and evolves the previous log into an over-complete face-log. Such a face-log can be used in *super-resolution* algorithms to obtain a high-resolution face image from the person in the video sequence. Having a video sequence as input, the proposed system in this thesis detects face images, and then uses the face quality assessment system to construct the aforementioned face-logs. Many different techniques have been implemented for face detection, facial feature extraction, face quality assessment, best, complete and over-complete face-log generation and finally super-resolution. Testing different parts of the system using more than 10 local and public databases produces good results.

TO ALL MY LOVED ONES

MARYAM FOR HER LOVE, MY PARENTS FOR THEIR SACRIFICES, AND MY SISTERS & BROTHERS FOR THEIR SUPPORTS!

PREFACE AND ACKNOWLEDGEMENT

This thesis is submitted in partial fulfillment of the requirements for the Doctor of Philosophy in Electrical Engineering at Department of Architecture, Design and Media Technology, Aalborg University.

The funding of this three-year Ph.D. study (Nov. 2007–Nov. 2010) came from "*Big Brother is Watching You!*" project. This project was funded by Danish National Research Council (FTP 274-07-0264).

I would like to thank all those who supported me during my Ph.D. study at Aalborg University. Special thanks go to my supervisor, Prof. Thomas B. Moeslund, who made me a researcher by his supports, his friendly supervision, and his great comments!
I am thankful for all the members of Computer Vision & Media Technology Lab from the secretaries to the head who all accepted me warmly and kindly from the beginning of my stay at the department.

Kamal Nasrollahi

November 2010,
Aalborg, Denmark

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1 Chapter 1: Introduction

1.1 Introduction

Nowadays, Biometric Recognition, or simply Biometrics, which refers to automatic recognition of individuals based on their physiological and/or behavioral characteristics, is a prominent field of research. Among all the biometrics like: face, fingerprint, hand geometry, iris, signature, DNA etc. face has an outstanding importance. Especially because of its contactless property, applications using this biometric are widely useful in surveillance cameras in public places like airports, banks, train stations, etc. Cameras providing inputs for such a system in public places are usually working constantly. Continues recordings from these cameras produce huge amount of video data. Considering a person passing by such a surveillance camera, a sequence of images of that person is captured by the camera. Faces in such videos can be detected in real-time but using all of the detected faces in almost any computer vision application is extremely demanding. Many of the detected faces are useless due to problems like not facing the camera, blurriness of the face, darkness of the face, small size of the face, and changes in the status of facial components (like closeness of the eyes, openness of the mouth). Using such face images would produce erroneous results in almost any facial analysis system. Furthermore, there are many faces that resemble each other very closely and keeping only some of them may suffice. Therefore, it is reasonable and necessary to use a mechanism for assessing the quality of the face images, i.e. face quality assessment [1]. This mechanism should discard useless face images and summarize the input video sequence to smaller sets containing some of the most expressive images of the video. These summarized sets, containing the most expressive images, are denoted Face-Log(s) [2, 3].

The face quality assessment term was first introduced by Griffin [1] where a face image is evaluated using some important features of the face. Face quality assessment in still images has been studied previously [4-9]. Kalka et al. [4] applied the quality assessment metrics originally proposed for iris to face images. Subasic et al. [5] presents a system to validate face images for using in identification documents. Such images should allow automatic face recognition to be successfully performed. The set of rules regarding the face image parameters that they use in their system is defined by the International Civil Aviation Organization (ICAO) [6]. ICAO defines thresholds and allowed ranges for parameters of the face image. Xiufeng et al. [7] presents an approach for standardization of facial image quality, and develops facial symmetry based methods for its assessment by which facial asymmetries caused by non-frontal lighting and improper facial pose can be measured. Fronthaler et al. [8] study orientation tensor with a set of symmetry descriptors to assess the quality of face images. Zamani et al. [9] try to deal with problems like shadows, hotspots, video artifacts, salt & pepper noise and movement blurring in the image and improve the quality of the image for the purpose of face detection.

Face quality assessment in video sequences for the purpose of constructing face-logs is relatively a new field of research [2, 11]. Fourney and Laganiere [2] after detecting and tracking faces using [10], extract six features for each face in each frame. Then, they assign a

score to each feature and after weighting these scores combine them into a general score. They use these general scores to construct face-logs by only keeping the face images with the highest scores. Xiong et al. [11] developed a metric based on bilateral symmetry, color, resolution and expected ratio (frontal face geometry) to determine whether a detected face image in a given surveillance video is suitable to be added to an on-the-fly database.

The decision of the face quality assessment system about an image in the original long video sequence to be in the face-log or not depends on the application. If the face-log is going to be used for example in video indexing [12], keeping the best face image of the sequence in the face-log (which is usually a frontal face) is sufficient. If the face-log is going to be used for recognition [13] it is a good idea to keep the best side view faces (if any) as well as the best frontal face image. If the face-log is going to be used in, for example a super-resolution system [14-27, to name just a few], it should contain some more images.

Super-resolution algorithms are used to obtain one or more high-resolution input image(s) from one or more low-resolution images. These algorithms are generally classified into two groups: reconstruction-based [14-23] and recognition-based [23-27]. Having usually one input low-resolution input, recognition-based super-resolution algorithms try to hallucinate the missing high-resolution details of the input image. Reconstruction-based super-resolution algorithms usually need more than one input low-resolution images. The inputs to these algorithms should be from the same scene or object and at the same time the inputs should have sub-pixel misalignments with each other. Therefore, if the face-logs generated by the face quality assessment are going to be used as the inputs to these kinds of algorithms, there is a need for some refinement, as it will be discussed later.

The proposed system in this thesis is a flexible face quality assessment system for summarizing high or low-resolution long video sequences for different purposes. The results of these summarizations (face-logs) are used in different real-world applications. The block diagram of the proposed system is shown in Figure 1-1. Such a system has four main blocks: Detection, Feature Extraction, Scoring, and Face-Log Generation. Having a video sequence as the input to the system, in the detection block, the faces and facial components are detected. Then, in the feature extraction block, the facial features that are used as quality measures for quality assessment are extracted for each face. Thereafter, in the scoring step of the quality assessment, the extracted quality measures are first normalized and then are combined to obtain a quality score for each face. Finally, depending on the application that is going to use the summarized results, different techniques are used to choose the best face images for the summarized result.

The motivation of this work is the lack of a fast, robust and automatic face quality assessment system working with faces of any sizes especially small ones which are common in video sequences from for examples surveillance cameras.

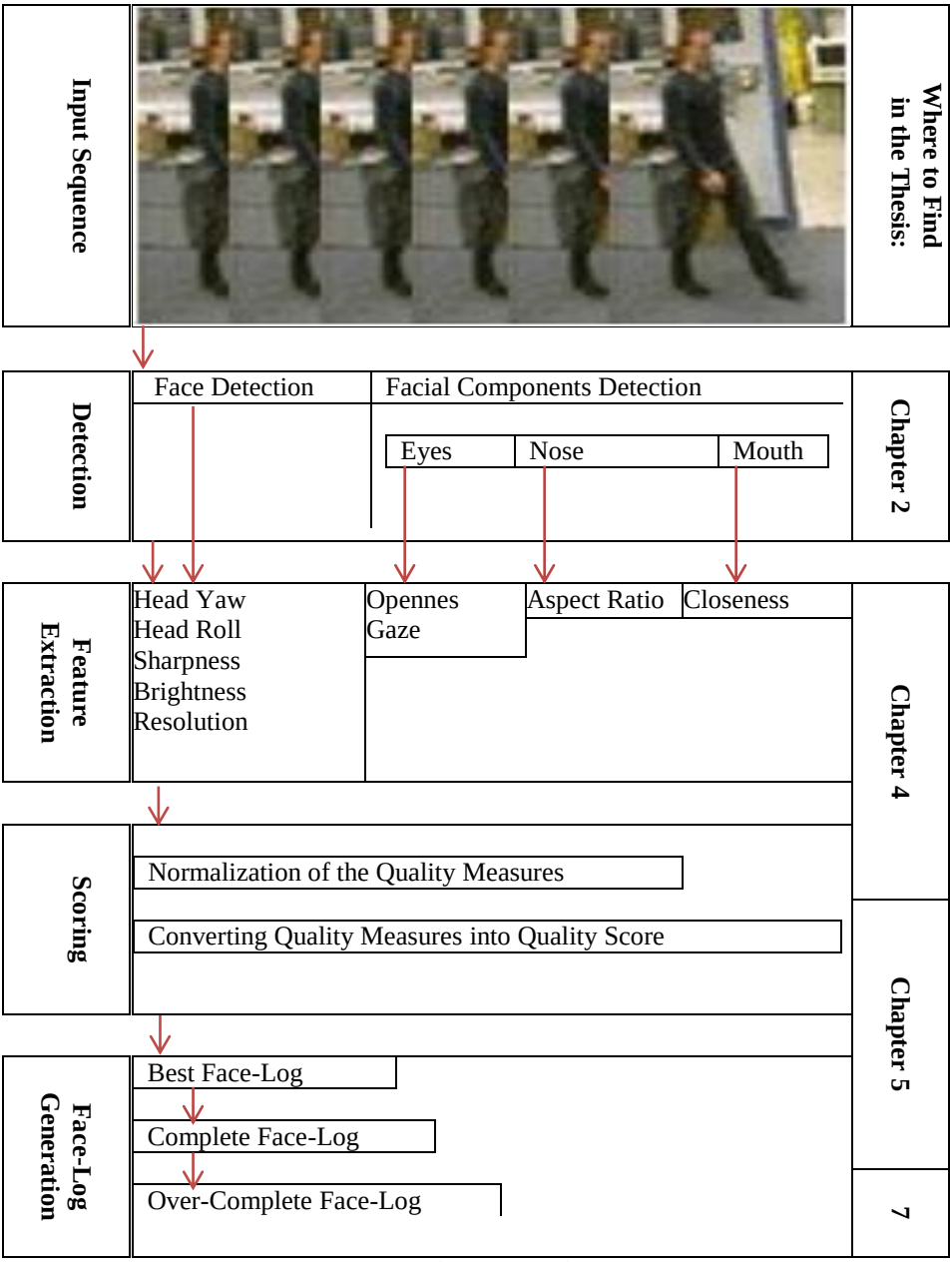


Figure 1-1: The block diagram of the proposed face quality assessment system

1.2 The Outline of the Thesis

During the completion of this thesis, several algorithms and possible methods are employed for the different parts of this system. Having a face detector with high detection rate and speed is still a challenging task in image processing and computer vision applications. In the detection

part, a number of different classifiers for face detection including fuzzy algorithms, neural networks, and genetic algorithms are investigated. The result is a face detector that combines these different classifiers. This algorithm and its comparison with the state-of-the-art Viola & Jones face detector [28] are discussed in chapter 2 of the thesis.

In chapter 3, the importance of the quality assessment, different types of quality measures, and the definitions of some of the most important quality measures are explained. Furthermore, the dynamic and static environments for quality assessment and their influence on the universe of discourse are also discussed in this chapter.

Chapter 4 of the thesis is devoted to the second block and the first part of the third block of the proposed system, i.e. feature extraction and normalization of the quality measures. In the feature extraction block the number and type of the facial features (that are used as quality measures) and their extraction methods are important. Several facial features are studied in the feature extraction block, and several systems are proposed using different types and numbers of facial features for different applications.

The methods for combining the normalized quality scores (second part of the third block) and subsequently the different techniques that are developed for face-log generation (fourth block of the proposed system) are discussed in chapter 5. It is shown that face-log generation can be done in many different ways and making the decision about the contents of the face-logs depends on the application that is going to use them afterwards. Two different possibilities for the contents of the face-log are investigated in this chapter. The first face-log contains the best face image of the sequence and can be used in applications like indexing the long video sequences that contain face images or in face recognition applications applied to long video sequences. The second face-log discussed in chapter 5, is a complete face-log. Such a face-log contains the best frontal and side-view faces of a long video sequence and can be considered as a complete and concise [2] representation of the video sequence.

In addition to the two above mentioned types one more kind of face-log has been studied in this thesis. This is face-logs for super-resolution algorithms. However, before going into the details of these kinds of face-logs in chapter 7, a survey on different methods for super-resolution in the literature has been performed in chapter 6.

Finally, the thesis is concluded in chapter 8. The motivations for further improvements applications of the current system in the future are also given in this chapter.

1.3 The Employed Databases

Having the general block diagram of Figure 1-1, several systems have been proposed for different parts of this block diagram. For testing these different systems 12 different databases are used. Three of these databases are prepared locally and nine of them are publicly available. These databases are explained in the text when they are being used.

1.4 Summary of the Contributions

Introducing a new cascaded classifier for face detection, developing easy and reliable methods for facial feature extraction, developing several systems for face quality assessment, proposing applicable methods for face-log generation, using the generated face-logs in real-world computer vision applications like face recognition and super-resolutions and introducing new hybrid super-resolution algorithms are among the contributions of this thesis. Following is a list of the papers that have been published during the accomplishment of the current Ph.D. thesis:

1.4.1 Journals and Book Chapters (Peer-Reviewed)

5. Kamal Nasrollahi, Thomas B. Moeslund, ***Extracting a Good Quality Frontal Face Image from a Low-Resolution Video Sequence***, under review, IEEE Transactions on Circuits and Systems for Video Technology.

4. Kamal Nasrollahi, Thomas B. Moeslund, Mohammad Rahmati, ***Summarization of Surveillance Video Sequences Using Face Quality Assessment***, To appear in International Journal of Image and Graphics, January 2011.

3. Kamal Nasrollahi, Thomas B. Moeslund, ***Complete Face-logs for Video Sequences Using Quality Face Measures***, IET International Journal on Signal Processing, vol. 3, nr. 4, pp. 289-300, 2009.

2. Kamal Nasrollahi, Mohammad Rahmati, Thomas B. Moeslund, ***A Neural Network-based Cascaded Classifier for Face Detection in Color Images with Complex Background***, A. Campilhom and M. Kamel (eds.) Image Analysis and Recognition, Springer Lecture Notes in Computer Science, vol. 51112, pp. 969-976, Springer Verlag Berlin Heidelberg, 2008.

1. Kamal Nasrollahi, Thomas B. Moeslund, ***Face Quality Assessment System in Video Sequences***, B. Schouten, N.C. Juul, A. Drygajlo and M. Tistarelli.(eds.) Biometrics and Identity Management, Springer Lecture Notes in Computer Science , vol. 5372, pp. 10-18, Springer Verlag Berlin Heidelberg, 2008.

1.4.2 Conference Proceedings (Peer-Reviewed)

6. Kamal Nasrollahi, Thomas B. Moeslund, ***Hallucination of Super-Resolved Face Images***, IEEE 10th International Conference on Signal Processing (ICSP), Beijing, China, 2010.

5. Kamal Nasrollahi, Thomas B. Moeslund, ***Finding and Improving the Key-Frames of Long Video Sequences for Face Recognition***, IEEE 4th International Conference on Biometrics Theory, Applications and Systems (BTAS), Washington DC, USA, 2010.

4. Kamal Nasrollahi, Thomas B. Moeslund, ***Hybrid Super-resolution using Refined Face-logs***, IEEE 2nd International Conference on Image Processing Theory, Tools and Applications (IPTA), Paris, France, 2010 (**Best Paper Award**).

3. Kamal Nasrollahi, Thomas B. Moeslund, ***Face-log Generation for Super-Resolution Using Local Maxima in the Quality Curve***, International Conference on Computer Vision Theory and Applications (VISAPP), Angers, France, 2010.

2. Kamal Nasrollahi, Thomas B. Moeslund, ***Real Time Face Quality Assessment for Face-log Generation***, International Conference on Machine Vision, Image Processing, and Pattern Analysis (MVIPPA), Bangkok, Thailand, 2009.

1. Kamal Nasrollahi, Thomas B. Moeslund, ***Face Quality Assessment System in Video Sequences***, the 1st COST 2101 Workshop on Biometrics and Identity Management (BIOID 2008), Roskilde, Denmark, 2008.

1.4.3 Conference Proceeding (NOT Peer-Reviewed)

1. Kamal Nasrollahi, Thomas B. Moeslund, ***Face Image Quality and its Improvement in a Face Detection System***, The 16th Danish Conference on Pattern Recognition and Image Analysis (DSAGM 2008), Copenhagen, Denmark, 2008.

References

[1] P. Griffin, "*Understanding the Face Image Format Standards*," American National Standards Institute/National Institute of Standards and Technology Workshop, Gaithersburg, Maryland, USA, 2005.

[2] A. Fourney and R. Laganiere, "*Constructing Face Image Logs that are Both Complete and Concise*," 4th IEEE Canadian International Conference on Computer Vision and Robot Vision, Canada, 2007.

[3] K. Nasrollahi and T.B. Moeslund, "*Face quality assessment system in video sequences*," 1st European Workshop on Biometrics and Identity Management, Denmark, 2008.

[4] N. Kalka, J. Zuo, N.A. Schmid, and B. Cukic, "*Image Quality Assessment for Iris Biometric*," SIPE Symposium on Defense and Security, International Conference on Human Identification Technology, Florida, USA, 2006.

[5] M. Subasic, S. Loncaric, T. Petkovic, and H. Bogunvoic, "*Face Image Validation System*," 4th International Symposium on Image and Signal Processing and Analysis, Zagreb, Croatia, 2005.

[6] ICAO Doc 9303, Parts I, II, III, "*Machine Readable Travel Documents specifications*," http://www.icao.int/ accessed December 2010.

[7] G. Xiufeng, Z. Stan, L. Rong, and P. Zhang, "*Standardization of Face Image Sample Quality*," International Conference on Advances in Biometrics, Seoul, Korea, 2007.
[8] H. Fronthaler, K. Kollreider, and J. Bigun, "*Automatic Image Quality Assessment with Application in Biometrics*," IEEE Conference on Computer Vision and Pattern Recognition, New York, USA, 2006.
[9] A.N. Zamani, M.K. Awang, N. Omar, and S.A. Nazeer, "*Image Quality Assessments and Restoration for Face Detection and Recognition System Images*," IEEE 2nd International Conference on Modeling & Simulation, Kuala Lumpur, Malaysia, 2008.
[10] Zhao, H. X., Huang, Y. S.: 'Real-Time Multiple-Person Tracking System'. In: IEEE 16th International Conference on Pattern Recognition, Quebec, Canada, 2002.
[11] Q. Xiong, and C. Jaynes, "*Mugshot Database Acquisition in Video Surveillance Networks Using Incremental Auto-Clustering Quality Measures*," IEEE Conferemce on Advanced Video and Signal-based Surveillance, Miami, USA, 2003.
[12] S. Eickeler, S. Muller, and G. Rigoll, "*Video Indexing Using Face Detection and Face Recognition Methods*," Workshop on Image Analysis for Multimedia Interactive Service, pp. 37-40, 1999.
[13] R. Chellappa, C.L. Wilson, and S. Sirohey, "*Human and Machine Recognition of Faces: A Survey*," Proceedings of the IEEE, vol. 83, no. 5, pp. 705-741, 2002.
[14] M. Irani, and S. Peleg, "*Improving Resolution by Image Registration*," Graphical Models and Image Processing, vol. 53, no. 3, 1999.
[15] M. Elad, and A. Feuer, "*Super-Resolution Reconstruction of Image Sequences*," IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 21, no. 9. pp. 817-834, 1999.
[16] A. Zomet, and S. Peleg, "*Super-Resolution from Multiple Images Having Arbitrary Mutual Motion*," In: S. Chaudhuri, Editor, Super-Resolution Imaging, Kluwer Academic, Norwell, pp. 195–209, 2001.
[17] S. Chaudhuri, "*Super-Resolution Imaging*," Kluwer Academic Publishers, 2nd edition, New York, 2002.
[18] M.E. Tipping and C.M. Bishop, "*Bayesian Image Super-Resolution*," Advances in Neural Information Processing Systems, vol. 15, pp. 1303-1310, 2002.
[19] S. Chaudhuri, and M.V. Joshi, "*Motion Free Super-Resolution*," Springer Science, New York, 2005.
[20] X. Li, Y. Hu, X. Gao, D. Tao and B. Ning, "*A multi-frame image super-resolution method*," Signal Processing, vol. 90, no. 2, pp. 405-414, 2010.
[21] L. Zhang, H. Zhang, H. Shen and P. Li, "*A Super-resolution reconstruction algorithm for surveillance images*," Signal Processing, vol. 90, no. 3, pp. 848-859, 2010.
[22] S. Baker, and T. Kanade, "*Limits on Super-Resolution and How to Break Them*," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 9, 2000.
[23] S. Baker and T. Kanade, "*Hallucinating Faces*," 4th IEEE International Conference on Automatic Face and Gesture Recognition, pp. 83-88, France, 2000.
[24] X. Wang and X. Tang, "*Face Hallucination and Recognition*," 4th International Conference on Audio- and Video-based Personal Authentication, IAPR, pp. 486-494U.K., 2003.
[25] G. Dedeoglu, T. Kanade, and J. August, "*High-Zoom Video Hallucination by Exploiting Spatio-Temporal Regularities*," IEEE International Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 151-158, USA, 2004.
[26] C. Su, Y. Zhuang, L. Huang, and F. Wu, "*Steerable Pyramid-based Face Hallucination*," Pattern Recognition, vol. 38, pp. 813-824, 2005.

[27] C. Miravet and F.B. Rodriguez, "*A Two-Step Neural-Network-based Algorithm for Fast Image Super-Resolution*," Image and Vision Computing, vol. 25, no. 9, pp. 1449-1473, 2007.
[28] P. Viola, M. Jones, "*Rapid Object Detection Using a Boosted Cascade of Simple Features*," IEEE Computer Vision and Pattern Recognition, pp. 511-518, 2001.

CHAPTER 2

FACE DETECTION

2 Chapter 2: Face Detection

2.1 Introduction

This chapter of the thesis is devoted to the first block of the proposed system shown in Figure 1-1, i.e. Face Detection. Face Detection is the process of finding and localizing faces inside a given image. In applications related to human face processing, such as face recognition, facial expression recognition, head-pose estimation, human-computer interaction, etc. face detection, has a considerable role. Most of these applications assume that the size and the position of the face(s) are known in the image. In real world applications, though, this is not the case and methods for estimating the position and the size of the face(s) are required. The success of the application that uses the detected faces depends highly on the accuracy of the face detection part. Therefore, it is of great importance to develop a face detector with a high detection rate in a reasonable time and at the same time keeping low the false positive rate (The regions that are detected as a face but are not a face are considered as false positives).

The rest of this chapter is organized as follows: the challenges of the modern face detection systems are discussed in the next section. Section 3 provides a survey on the different types of face detection algorithms. Section 4 describes the details of our proposed face detection system and section 5 concludes the chapter.

2.2 Challenges

Even though face detection has been studied for a long time, the researchers are still trying to develop a more advanced face detector. This is due to the challenges that can fail even the more recent face detectors. These challenges can be listed as follows [1, 2]:

- Facial expression: changes in the facial expression can directly affect the shape, size and structure of facial components and thus the face appearance (Figure 2-1(a)).
- Head-pose: most of the face detectors use the features of the facial components like the relative size and position of the components. Extracting these kinds of features may not be easy when the head is rotated. The situation becomes worse when some or all of the facial components are hidden due to the changes in the head-pose (Figure 2-1(b)).
- Presence or absence of structural elements: facial features such as beards, mustaches, hair and glasses may or may not be present and there is a wide range of changes among these components including shape, color, and size (Figure 2-1(c)).
- Occlusion: faces may be partially occluded by other objects. In an image with a group of people, some faces may partially occlude other faces (Figure 2-1(d)) or by some objects in the scene (Figure 2-1(e)).
- Image orientation: face images directly vary for different rotations about the camera's optical axis (Figure 2-1(f)).

- Imaging condition: when the image is formed, factors such as lighting (spectra, source distribution and intensity) and camera characteristics (sensor response, lenses) affect the appearance of a face (Figure 2-1(g)).

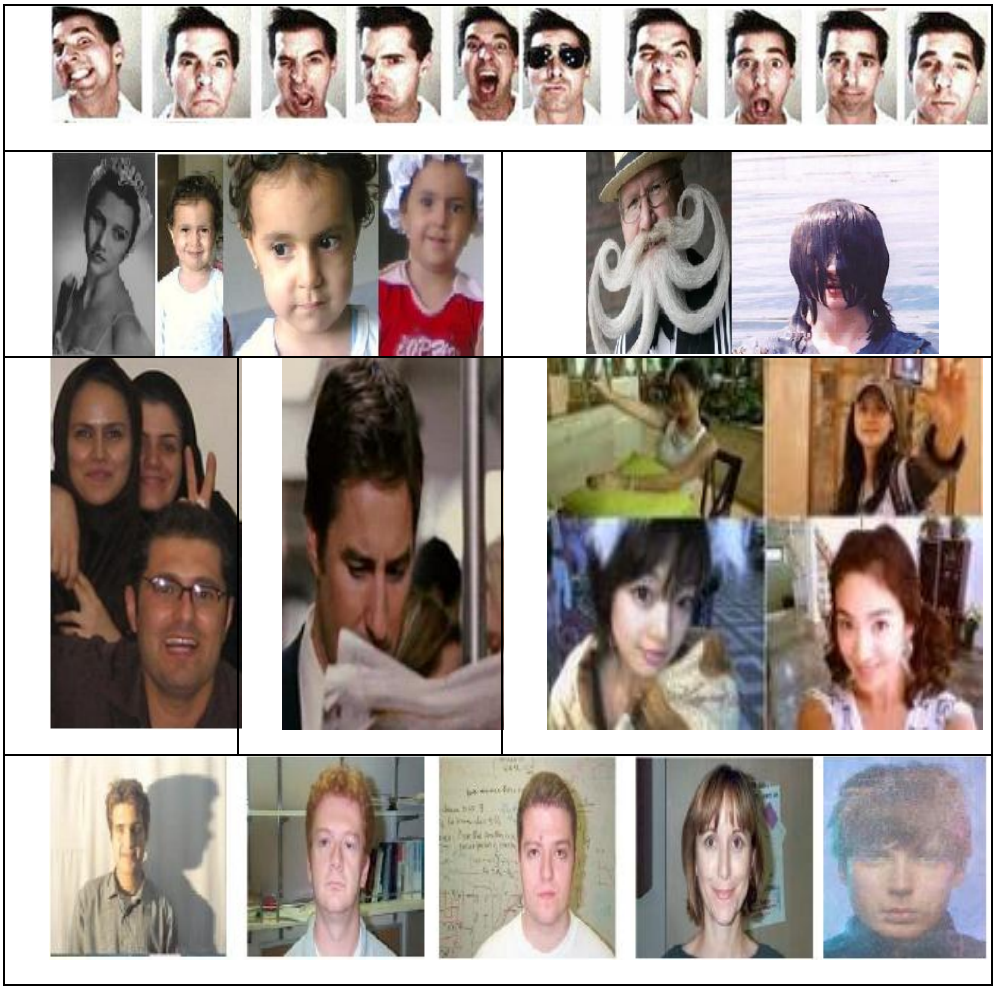


Figure 2-1: Challenges of face detection systems

a		
b		c
d	e	f
G		

2.3 Literature Survey

There are few papers [1, 2] providing surveys on the huge number of papers published on face detection in the literature. These surveys classify different face detection methods into different classes. In order to understand the contributions of the proposed system, in this section, we review the different classes of face detection algorithms and their pros and cons.

2.3.1 Knowledge-based Methods

Rule based methods encode human knowledge of what constitutes a typical face. The rules are derived from the researcher's knowledge of human faces. Usually, the rules capture the relationships between facial features [1, 2].

It is easy to come up with simple rules to describe the features of a face and their relationships. For example, a face often appears in an image with two eyes that are symmetric to each other, a nose, and a mouth. One problem with this approach is the difficulty in translating human knowledge into well-defined rules. If the rules are detailed and strict, they may fail to detect faces that do not pass all the rules. If the rules are too general, they may give many false positives. Moreover, it is difficult to extend this approach to detect faces in different poses since it is challenging to enumerate all possible cases. On the other hand, heuristics about faces work well in detecting frontal faces in uncluttered scenes [1, 2].

2.3.2 Feature Invariant Methods

These algorithms try to find the structural features that exist even when the pose, viewpoint, or lighting conditions vary, and then use these features to locate faces [1, 2]. Typical features are facial features, texture of the face, skin color and combination of multiple features.

Usually in this method, the facial features are easily extracted using edge detectors. One problem with these feature based algorithms is that the image features can be severely corrupted due to illumination, noise, and occlusion. Feature boundaries can be weakened for faces, while shadows can cause numerous strong edges which together render perceptual grouping algorithms useless [1, 2].

2.3.3 Template matching methods

In this method usually several standard patterns of a face are stored to describe the face as a whole or the facial features separately. The correlations between an input image and the stored patterns are computed for detection. These templates can be either predefined face templates like shape templates or deformable templates like active shape models. The predefined templates can be defined manually or by a parameterized function [1, 2].

This approach can be implemented easily. However, it has proven to be inadequate for face detection since it cannot effectively deal with variation in scale, pose, and shape. Multi-resolution, multi-scale, sub-templates, and deformable templates have subsequently been proposed to achieve scale and shape invariance [1, 2].

2.3.4 Appearance-based Methods

In contrast to the template matching methods where templates are predefined by experts, the templates in appearance-based methods are learned from examples in images. In general, appearance-based methods rely on techniques from statistical analysis and machine learning to find the relevant characteristics of face and non-face images. The learned characteristics are in the form of distribution models or discriminate functions that are consequently used for face detection. Meanwhile, dimensionality reduction is usually carried out for the sake of computation efficiency and detection efficacy [1, 2].

Some appearance-based methods are: eigenface methods, distribution-based methods, neural networks, support vector machines, hidden Markov models, etc.

2.4 The Proposed Method

In general there are two issues in face detection systems: detection rate and speed. They often work against each other. Different proposed applications in the literature have tried to find an acceptable tradeoff between these two.

The systems proposed in [3-7] are examples of systems optimized with respect to speed. Gottumukall and Asari [3] reported a face detection system capable of detecting faces in real time from a streaming color video. In their system after a skin color segmentation the principal component analysis is used to determine if a particular skin region is a face or not. Viola and Jones [4] introduce AdaBoost with a cascade scheme and apply an integral image concept for face detection. They propose a two-class AdaBoost learning algorithm for training efficient classifiers and a cascade structure for rejecting none face images. Wu et al. [5] propose an efficient face candidates selector for face detection tasks in still gray level images. They are discovering the eye-analogue segments at a given scale by finding regions which are roughly as large as real eyes and are darker than their neighborhoods. Then a pair of eye-analogue segments are hypothesized to be eyes in a face and combined into a face candidate if their placement is consistent with the anthropological characteristic of human eyes. Cheng et al. [6] after compensating the colors of the input images and deskewing tilted faces locate mouth corners and determine a discriminate function for positioning eyes. Zhong et al. [7] use a luminance-conditional distribution for modeling the skin color information and then by morphological operations extract the skin-region rectangles. Finally, they use template matching based on a linear transformation to detect a face in each skin-region rectangle.

Neural networks are well known classifiers which have been used widely in face detection [8-13] when detection rate is in focus. Rowley et al. [8] present a neural network based upright frontal face detection system. A retinally connected neural network examines small windows of an image, and decides whether each window contains a face. Mansour et al. [9] propose a face detection algorithm based on light control, skin detection and color segmentation techniques. Their method detects the face rectangles that contain eyes and mouth and then constructs expected regions resulted from skin detection and color segmentation stages. Next they search inside them for any possible face features (eyes, and mouth) and pass these expected mouth and eyes rectangle to a neural network to confirm the presence of a face.

However, high computations between the layers of the neural networks and also problems in adjusting the topology of the network limit their usage. These problems may become more complicated in some cases. For example, computation will be higher if the network scans the entire image for finding possible faces without any prior knowledge [8, 12].

The proposed face detection system in this thesis (Our first proposed system) provides solutions to the above mentioned problems using a neural network. First problem, reducing the computations, is dealt with by the use of a pre-classifier which reduces the number of regions which are given to the neural network. This decreases the computations greatly. This pre-classifier has considerable importance because if it makes decisions based on a set of complicated and concrete rules then it is possible to miss some of the faces in the image. As a compromise we have developed a fuzzy inference engine which by a set of flexible rules, over a small number of reliable and easy to extract features, eliminates some of the non-face regions, hence reduce the number of regions presented to the neural network. Second problem, adjusting the network topology, is dealt with by applying a genetic algorithm to find the best network parameters in the context of the face detection problem.

The block diagram of our proposed system is shown in the Figure 2-2. Given a color image as the input to the system, the search space is reduced by segmenting skin (or skin like) regions from the non-skin ones. For each segmented region a cascaded classifier including a pre-classifier and a main classifier determines whether this region contains a face or not. In the pre-classification step a small number of reliable and easy to extract features are computed and fed to a fuzzy inference engine. If the result of the fuzzy inference engine shows the presence of a face in the current region, this region will be sent to the main classifier for making the finial decision. The output of the system will show the detected face(s) in the given image. The following subsections will discuss the segmentation method, the classification method and the experimental results.

2.4.1 Segmentation

Given a color image as input to the proposed system, the first step is to reduce the search space. Using a Bayesian approach similar to [14-16] we separate skin or skin-like regions from the non-skin ones.

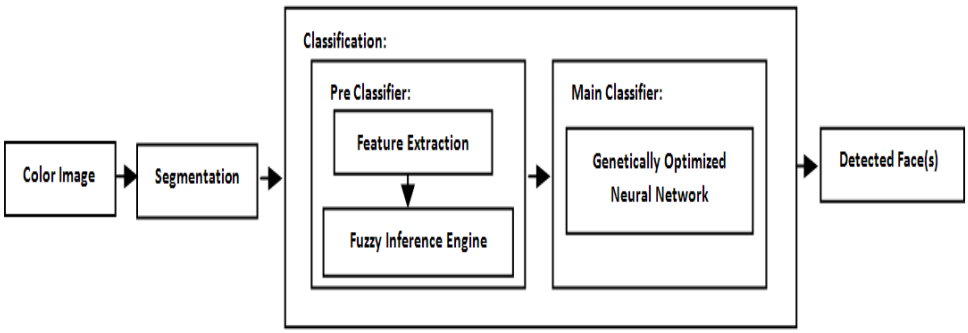


Figure 2-2: The block diagram of the proposed face detection system.

In order to model skin color [14-17] we have selected about 50,000 skin samples from randomly chosen people of different ethnics from the web. We use a Chromatic color space, in which the illumination is eliminated by a normalization process using the following equation:

$$r = \frac{R}{R+G+B}, g = \frac{g}{R+G+B}, b = \frac{B}{R+G+B} \qquad 2\text{-}1$$

Figure 2-3 (left) shows the distribution of the skin samples in the chromatic color space. This distribution can be modeled by a Gaussian distribution (Figure 2-3 (right)) with the following parameters:

$$m = E(x), c = E\left((x,m)(x,m)^T\right), \quad x = (r,b)^T \qquad 2\text{-}2$$

Where m is the mean, c is the covariance and x is a vector of the values of r and b. Using this Gaussian distribution it is possible to determine the similarity of each pixel of the input image to a skin pixel. Hence, given an input image in RGB color space it will be converted to the above Chromatic colors and then using the Mahalanobis distance (2-3), similarity to a skin pixel is calculated for each pixel.

$$S(r,b) = \exp[-0.5(x-m)^T c^{-1}(x-m)] \qquad 2\text{-}3$$

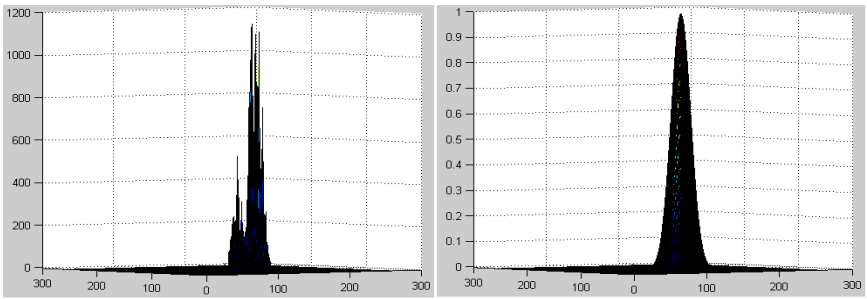


Figure 2-3: (Left) Distribution of skin samples in chromatic color space and (right) their Gaussian distribution model.

Since the value of the above equation for each pixel is between 0 and 1 the output of the above step will be a probability image (see Figure 2-4). Using an adaptive threshold [16] this image is segmented into a binary image. Figure 2-4 shows an input color image, its skin like image (probability image) and its segmented counterpart (binary image) as a result of the above process. It is obvious that in this binary image some of these regions are not faces but other parts of the body (naked parts of hands and the legs) and others are related to the things which have a skin like color (e.g. horizontal bar at the left of the image). So for making an initial decision about a region being a face or not we use a pre-classifier which is the matter of the next subsection.

2.4.2 Classification

The segmented image from the previous step has separated skin or skin-like regions from the non-skin ones. Now, the system goes through these skin or skin-like regions and classifies them as face or non-face. The main classifier is a neural network. The advantage of neural networks as classifier in the face detection problem is their auto ability in extraction the characteristics of the complicated face templates [9]. However, this auto ability costs high computations between the layers of the network. Furthermore, the topology of the network influences both the computation and achieving acceptable results.

In order to reduce the computation of the system we avoid the neural network from blindly scanning all skin or skin-like regions in the segmented image. Instead we use the neural network in a cascaded classifier. The cascaded classifier consists of two different classifiers. At the first level a fuzzy inference engine, which we call it a pre-classifier goes through the skin or skin-like regions and extracts some features, and at the second level the neural network will

deal with the regions, that successfully have passed the pre-classifier, to make final decision. As any other cascaded classifier ours is designed in such a way that the less computational classifier will be in the first step and the most computational one will appear last. The details of these classifiers are described in the following subsections.

Figure 2-4: (Left) input color image, (middle) probability image and (right) segmented image

2.4.2.1 Pre-classifier: Fuzzy Inference Engine

By scanning the binary image of the previous step from the top left to the bottom right each group of connected pixels is considered a region. Based on the extracted features for each segmented region the initial decision will be made by the pre-classifier, i.e. faces or not.

The features involved in the decision making are: number of holes inside a region, center and orientation of the region, length and width of the region, ratio of the length to the width, the ratio of the holes to their surrounding region, and correlation between the region and a face template (Figure 2-5). Note that some of these features are used to calculate some of the others.

Figure 2-5: Face Template used in calculating the correlation [16]

The reasons for choosing these features are their ease of extraction in terms of computation and their reliability when used together. Regarding the details of feature extraction, the reader is referred to [16].

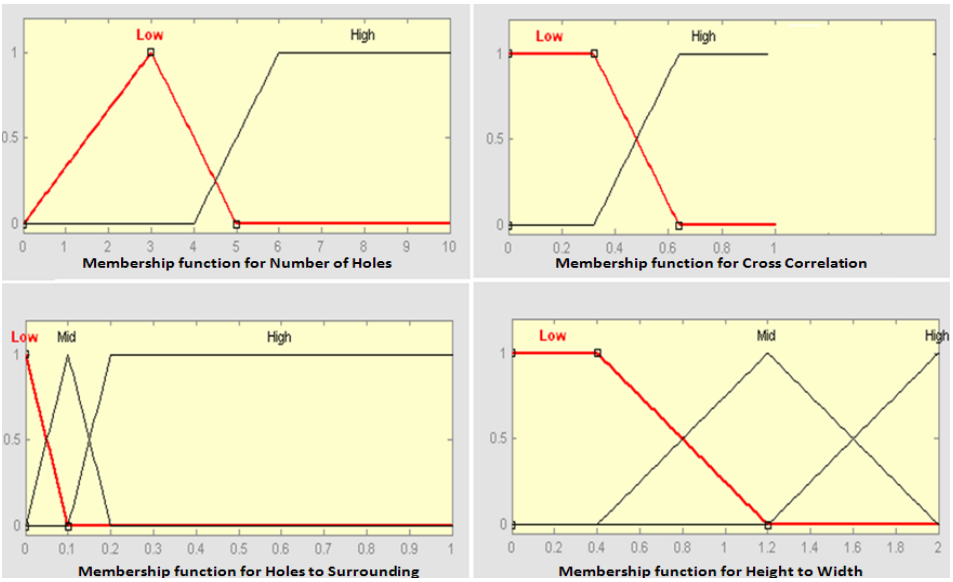


Figure 2-6: Designed member functions for fuzzy inference engine input variables

As the first step of the cascaded classifier a fuzzy inference engine accepts the extracted features for each region and using a set of fuzzy rules tries to make an initial decision about them being faces or not. We have used a Mamdani [18] fuzzy model for implementing this fuzzy inference engine. This engine has four inputs which correspond to the used features and they are Number of holes, correlation with a face template, the ratio of the area of the holes to their surrounding and the length of the region to its width. These inputs are fuzzified using the membership functions shown in Figure 2-6. Hereafter the rules shown in Figure 2-7 are applied. The weights of all rules are equal and are considered as one. The used aggregation method for the rules is Maximum value and the defuzzification method is the Mean value of maximum.

1. If **number of holes** is **high** then the output is **not a possible face**.
2. If **cross correlation** is **low** then the output **is not a possible face**.
3. If **holes area to their surrounding area** is **low** then the output is **not a possible face**.
4. If **holes area to their surrounding area** is **high** then the output is **not a possible face**.
5. If **height to width** is **low** then the output is **not a possible face**.
6. If **height to width** is **high** then the output is **not a possible face**.
7. If **number of holes** is **low** and **holes area to their surrounding area** is **mid** then the output is **a possible face**.
8. If **cross correlation** is **high** and **holes area to their surrounding area** is **mid** then the output is **a possible fac**
9. If **cross correlation** is **high** and **height to width** is **mid** then the output is **a possible face**.

Figure 2-7: The used rules in Fuzzy Inference Engine

Using this fuzzy inference engine has established an acceptable tradeoff between the computation and the missed faces while the rate of correct detection is acceptably high. Figure 2-7 shows how the output of the fuzzy inference engine changes with respect to the changes of its input parameters.

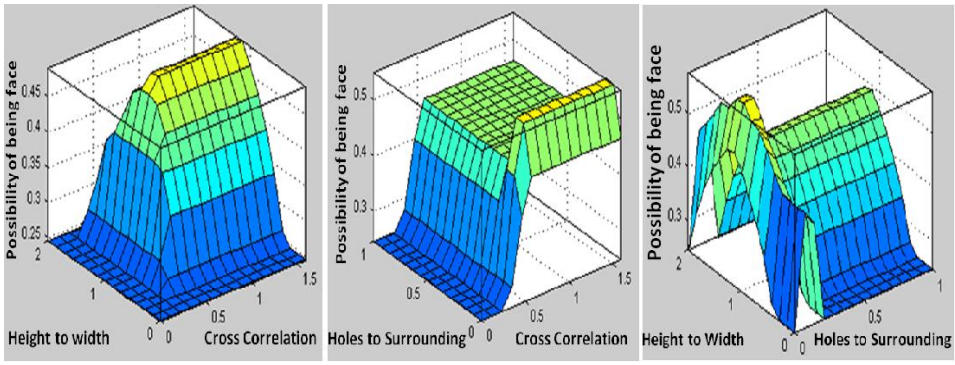


Figure 2-8: Change of the Fuzzy Inference Engine output with respect to change of its input

If the output of the above fuzzy inference engine indicates the presence of a face in a region, this region will be fed into the main classifier which is described in the following subsection.

2.4.2.2 Main classifier: Optimized Neural Network by a Genetic Algorithm

The accepted regions from the previous step are resized and then fed to our main classifier which is a neural network. Following the work done in [19] all the parameters involved in the topology of the network (e.g. number of neurons in hidden layer, activation function for each neuron, learning rates, connections weights, etc.) are optimized by a genetic algorithm.

According to Figure 2-9 each genotype codes a phenotype or candidate solution. The phenotypes (the resulting neural networks) are trained with the back-propagation algorithm. The evaluation of a phenotype determines the fitness of its corresponding genotype. The evolutionary procedure deals with genotypes, preferably selecting genotypes that code phenotypes with a high fitness, and reproducing them. Genetic operators are used to introduce variety into the population and to test variants of candidate solutions represented in the current population. In this way, over several generations, the population will gradually evolve towards genotypes that correspond to phenotypes with high fitness [19]. In our work the selection method, cross over method, possibility of cross over are roulette, one point, 0.9, respectively. This algorithm converges after 30 generation.

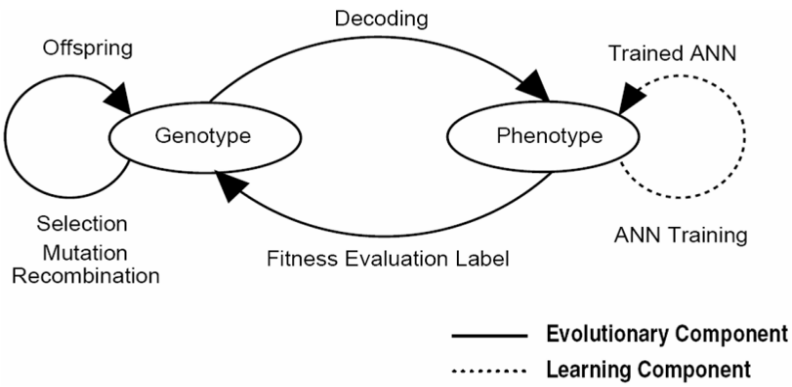


Figure 2-9: Design process of the evolutionary network topology [19]

The fitness of each phenotype is evaluated by calculating the sum of the squares error for its associated network. Figure 2-10 shows the classification error for the best network on the training and validation sets. 400 epochs is sufficient for training each network. In our work the selection method is roulette and the cross over method and its possibility are one point and 0.9, respectively. This algorithm converges after 30 generation.

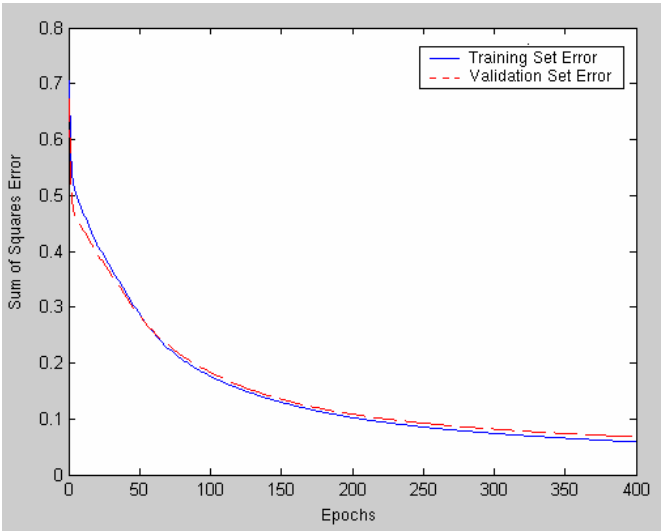


Figure 2-10: Networks classification error on the training and validation sets vs. epochs.

2.4.3 Experimental Results

We evaluate our system using two different databases. One of them (DB1) is taken from the California Institute of Technology (CIT) [21] and the other one (DB2) is provided manually. Most of these images have complex background and are taken in indoor and outdoor environment. The CIT database has 450 color images with the dimension of 320×240 pixels which come from 27 different persons in different conditions and at different scale. The manually provided database contains 100 different color images. The number of faces in the images of this database is varying from 1 to 25. The smallest face in this database is 21×25 and the biggest one is 252×258 pixels. Some of these images have obstacles such as glasses, beard, and moustache and make up. Some faces have overlaps with each other. Although most of the faces in these databases have frontal view, there are some which have soft in-plan rotations. Table 2-1 shows the results of testing our system on the above databases.

Table 2-1: The testing results of the proposed system

	Number of Faces	Correct Detections	False Positive	False Negative	Correction Rate
DB1	450	407	78	43	90.4%
DB2	220	201	49	19	91.3%

We compare our results with two other systems. The first one (System 1) is the proposed system in [22]. This is the system found in the literature which is most similar to ours. This system uses a modified genetic algorithm to find the best rules for the face detection problem in the rule-layer of a five layer neuro-fuzzy classifier. The second system (System 2) uses a

free open source computer vision library (OpenCV) containing a face detection function being proposed in [23]. This system uses simple Haar-like features and a cascade of boosted classifiers as its statistical model [4]. Table 2-2 shows the results of this comparison on the CIT database (database 1). It is obvious from this table that our system holds the best results in terms of detection rate and false detection1. While keeping the same detection rate as System 1 our system is ten times faster. The reasons for this difference in speed are: first, the proposed system in [22] is dealing with all pixels in the image while our system uses a cascaded classifier, which discards all irrelevant regions. Second, [22] is using a modified genetic algorithm for optimizing their system while we use the genetic algorithm for optimizing just the neural network. Furthermore it should be noticed that our system is much faster than [22] in training.

Table 2-2: The proposed system vs. the systems in [22] using CIT database (DB1).

	System 1	System 2	Our System
Detection rate	86%	90.1%	90.4%
False Detection	319	1088	121
Images in Training set	150 Face & Non Face	6000 Face & Non Face	200 Face & Non Face

Figure 2-11 illustrates some of the results of the proposed system. Below each image in this figure there is a set of numbers which are: Number of face(s) in the image, Number of corrected detection(s), Number of missed face(s) and Number of false positive(s), respectively. As it can be seen from these images the proposed system can detect different faces in different conditions. In image (a) the system is applied to a crowded scene in an outdoor environment. Although the system is working well for most of the faces in the images there are some faces that have not been detected. The problem with these faces in this image is that they cannot satisfy any of the rules in the fuzzy inference engine i.e. there is a lack of holes inside the faces (as a result of segmentation) and incorrect ratio between their width and length. The face in images (b and c) are indoor and under artificial lightening condition. The images (d, e, f, g and h) in different conditions show the faces in different statues. Faces in image (d and i) are detected even though hats are present. The images in part *j* and *k* are from the CIT database in different lightening conditions. In image (l) the faces of the two leftmost girls are detected correctly even though they overlap. In images (m and r) behind the detected faces there is a false positive which is due to the very small size of the skin regions. In image (o) you can see a correctly detected face despite of presence of an obstacle (newspaper).There are some detected faces in images (a, l, o, p and s) with obstacles. The faces in g, h, n and o have rotation. And finally it can be seen that the proposed system is working well even with images having a complex background.

Figure 2-11: Some of the results of our proposed system.

2.5 Summary

The proposed system in this chapter of the thesis deals with the face detection problem in color images with complex background using a cascaded classifier, combining neural networks (optimized by genetic algorithm) and fuzzy inference engine. After separating the skin (or skin like) regions from the non-skin ones, reliable and easy to compute features are extracted and used by a well-designed fuzzy inference engine with a set of flexible rules. Finally a neural network, with its topology optimized by the genetic algorithm, makes the final classification.

We test our system on two databases of color images. One of them is taken from California Institute of Technology and the other one is provided manually. The detection rates for these

databases are 90.4% and 91.3%, respectively. While this detection rate is comparable to state of the art neural network classifiers, the processing time (and also training time) is much faster. Using a more sophisticated segmentation method, adding some additional rules to the fuzzy inference engine and also weighting them can improve the performance of the face detector.

References

[1] M. Yang, D. Kriegman, and N. Ahuja, "*Detecting Faces in Images: A Survey*," IEEE Transaction Pattern Analysis and Machine Intelligence, vol. 24, no. 1, pp. 34-58, 2002.

[2] E. Hjelmas, and B.K. Low, "*Face Detection: A Survey*," Computer Vision and Image Understanding, vol. 83, pp. 236-274, 2001.

[3] R. Gottumukkal, and V.K. Asari, "*Real Time Face Detection From Color Video Stream-based on PCA Method*," Proceedings of 32nd Applied Imagery Pattern Recognition Workshop, pp. 146-150, 2003.

[4] P. Viola, M. Jones, "*Rapid Object Detection Using a Boosted Cascade of Simple Features*," IEEE Computer Vision and Pattern Recognition, pp. 511-518, 2001.

[5] W. Jianxin, Z. Zhi-Hua, "*Efficient Face Candidates Selector for Face Detection*," Pattern Recognition, vol. 36, no. 5, pp. 1175-1186, 2003.

[6] C. Chiang, and C. Huang, "*A Robust Method for Detecting Arbitrarily Tilted human Faces in Color Images*," Pattern Recognition, vol. 26, no. 16, pp. 2518-2536. 2005.

[7] Z. Jin, Z. Lou, J. Yang, and Q. Sun, "*Face detection using template matching and skin-color information*," Neurocomputing, , vol. 70, no. 4, pp. 794-800, 2007.

[8] H. Rowley, S. Baluja, and T. Kanade, "*Neural Network-Based Face Detection*," IEEE Transaction Pattern Analysis and Machine Intelligence, vol. 20, no. 1, pp. 23-38, 1998.

[9] I.M. Mansour, and S.R. Weshah, "*A Robust Algorithm for Face Detection in Color Images*," International Conference on Visualization, Imaging, and Image Processing. Spain, 2005.

[10] C. Garica, and M. Delakis, "*Convolutional Face Finder: a Neural Architecture of Fast and Robust Face Detection*," IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 26, no. 11, pp. 1408-1423, 2004.

[11] S.-H. Lin, S.-Y. Kung, and L.-J. Lin, "*Face Recognition/Detection by Probabilistic Decision-Based Neural Network*," IEEE Transaction on Neural Networks, vol. 8, no. 1, pp. 114-132, 1997.

[12] P. Juell, and R. Marsh, "*A Hierarchical Neural Network for Human Face Detection*," Pattern Recognition, vol. 29, no. 5, pp. 781-787, 1996.

[13] R. Feraud, O.J. Bernier, J.-E. Villet, and M. Collobert, "*A Fast and Accurate Face Detector-based on Neural Networks*," IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 22, no. 1, pp. 42-53. 2001.

[14] M.-H. Yang, and N. Ahuja, "*Detecting Human Faces in Color Images*," Proceedings of IEEE International Conference on Image Processing, vol. 1, pp. 127-139, 1998.

[15] H. Wu, Q. Chen, and M. Yachida, "*Face Detection from Color Image Using a Fuzzy Pattern Matching Method*," IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 21, no. 6, pp. 557-563, 1999.

[16] K. Nallaperumal, R. Subban, K. Krishnaveni, L, Fred, and R.K.Selvakumar, "*Human Face Detection in Color Images Using Skin Color and Template Matching Models for*

Multimedia on the Web," International Conference on Wireless and Optical Communications Networks, IFIP, 2006.
[17] A.M. Ferreira, and M.V. Correia, "Face Detection-based on Skin Color in Video Images with Dynamic Background," ICIAR, pp. 254-262, Canada, 2007.
[18] C. Giovanna, M. Anna, and C. Mencar, "*Design of Transparent Mamdani Fuzzy Inference Systems*," Computational Intelligence Laboratory, University of Bari, Bari, Italy, 2000.
[19] A. Fiszelew, P. Britos, A. Ochoa, H. Merlino, and E. Fernández, "*Finding Optimal Neural Network Architecture Using Genetic Algorithms*," Advances in Computer Science and Engineering, Research in Computing Science 27, pp. 15-24, 2007.
[20] http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html
[21] http://www.vision.caltech.edu/archive.html
[22] C. Lin, H. Chuang, and Y. Xu, "*Face Detection in Color Images using Efficient Genetic Algorithms*," Optical Engineering vol. 45, no. 4, pp. 1-12, 2006.
[23] G. Bradski, A. Kaehler, and V. Pisarevky, "*Learning-based Computer Vision with Intel's Open Source Computer Vision Library*," Intel Technol. vol. 9, no. 2, pp. 119-130 (2005)

CHAPTER 3

IMAGE QUALITY ASSESSMENT

3 Chapter 3: Image Quality Assessment

3.1 Introduction

This relatively short chapter gives an overview of the importance of image quality assessment and the quality measures that are used for this purpose. Image quality assessment is employed to assess the amount of the useful information that is present in a given image [1]. It is obvious that images with the higher qualities are preferred to their lower counterparts in most of the image processing and computer vision systems. Therefore, image quality assessment is of great importance in such systems.

The rest of this chapter is organized as follows: different types of quality measures are described in the next section. Section 3 compares the employment of image quality assessment in dynamic and static environments wherein the universe of discourse comes into account. Finally section 4 concludes this chapter.

3.2 Different Types of Quality Measures

Usually some of the features of a given image are used as quality measures in image quality assessment. It is expected that these measures follow the human perception of quality. Image quality metrics are of a significant importance in image processing applications as they can be used to adjusting the quality of a given image, judging about the quality of a given image, improving the results of image processing algorithms, benchmarking different systems and algorithms [2].

Quality measures are generally divided into two categories: subjective and objective [1, 2, and 3]. Following is a brief description of these measures.

3.2.1 Subjective Quality Measures

Subjective quality measures are based on the human perception of the quality of the images. In fact, the most reliable way of assessing the quality of an image is by subjective evaluation [2]. Subjective quality measures are usually used in the mean opinion score (MOS) way. In this method, few human observers are asked to assess the quality of a given image and then their MOS is calculated and used as the quality measure of the image.

3.2.2 Objective Quality Measures

These types of quality measures have a standard mathematical definition and they are generally divided into two groups: binary and unary measures [4, 5]. Binary measures are usually used when there is a reference image that the input image is compared against it. Thus, binary

measures assess the closeness or similarity of the input images to a reference image. Unary measures do not compare the input image with a reference image. Instead, they assign a numerical value to the input image based on the measurements of an image field [5]. Following subsections describe some of the most important binary and unary quality measures. A comprehensive list of these measures can be found in [4, 6]. Both types of objective (binary and unary) and subjective quality measures have been used throughout this thesis.

3.2.2.1 Binary Measures

Considering *I(i, j)* as a given input image and *R(i, j)* as the reference image, following is a list of binary measures that are frequently used in quality assessment. It is assumed that both images are of the same size and that is $M \times N$ pixels [3-5].

- L_p norms: It is defined as in 3-1:

$$L_p = \{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}|I(i,j) - R(i,j)|^p\}^{1/p} \quad 3\text{-}1$$

  where p can be 1, 2, or 3. L_p norms cannot take into account the dependency (ordering, pattern, etc.) between the signal samples [1, 2].

- Normalized cross correlation: Cross-correlation is a measure for determining the similarity between two images. This is also known as a sliding dot product or inner-product. For image-processing applications in which the brightness of the input image and reference image can vary due to lighting and exposure conditions, the images can be first normalized. The definition of the normalized cross correlation is as in 3-2:

$$NCC = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N} I(i,j)R(i,j)}{\sum_{i=1}^{M}\sum_{j=1}^{N}[I(i,j)]^2} \quad 3\text{-}2$$

- Average difference: It is defined as in 3-3:

$$AD = \sum_{i=1}^{M}\sum_{j=1}^{N}\frac{[I(i,j) - R(i,j)]}{MN} \quad 3\text{-}3$$

- Structural Contents: It is defined as in 3-4:

$$SC = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}[I(i,j)]^2}{\sum_{i=1}^{M}\sum_{j=1}^{N}[R(i,j)]^2} \quad 3\text{-}4$$

- Mean Square Error: The mean square error or MSE determines the difference between an input image and a reference image. MSE is a risk function, corresponding to the expected value of the squared error loss or quadratic loss. MSE measures the average of the square of the error. It is defined as in 3-5:

$$MSE = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}[I(i,j) - R(i,j)]^2}{MN} \quad 3\text{-}5$$

- Correlation Coefficient: Correlation coefficient finds the relationship between an input image and a reference image in terms of the least squares [7]. It is defined as in 3-6:

$$CC = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}[(I(i,j),-avg(I))(R(i,j),-avg(R))]}{\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}[I(i,j),-avg(I)]}\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}[R(i,j),-avg(R)]}} \quad 3\text{-}6$$

- Structural Similarity: This measure takes into account contrast, luminance and the structure of a given image and a reference image to determine their similarity [7]. It is defined as in 3-7:

$$SS = avg[\frac{(2avg(x)avg(y)+C_1)(2\tau(x,y)+C_2)}{(avg(x)^2+avg(y)^2+C_1)(\sigma(x)^2+\sigma(y)^2+C_2)}] \quad 3\text{-}7$$

where C_1 and C_2 are constants, x and y are sub-images of the input image I and the reference image R, respectively, σ returns the standard deviation of its parameter and τ is defined as in 3-8:

$$\tau(x,y) = \frac{1}{MN-1}\sum_{i=1}^{M}\sum_{j=1}^{N}[(x(i,j),-avg(x))(y(i,j),-avg(y))] \quad 3\text{-}8$$

The maximum values for both correlation coefficient and structural similarity measures are one, which indicate the perfect correlation and similarity, respectively, between the input image I and the reference image R.

3.2.2.2 Unary Measures

Unary measures can be generally expressed by 3-9 in which O is an operator on the input image I of size $M \times N$ pixels [1, 2]:

$$Um = \sum_{i=1}^{M} \sum_{j=1}^{N} O\{I(i, j)\} \quad \text{3-9}$$

If z_i represents a random variable indicating intensity in an image, $p(z)$ the histogram of the intensity levels in a region, L the number of possible intensity levels, then a sample list of unary measures which calculate image quality based on textural content are discussed below [3-5]:

- Mean: Mean is the measure of average intensity in an image. It is defined as in 3-10:

$$m = \sum_{i=1}^{L} z_i \, p(z_i) \quad \text{3-10}$$

- n^{th} moment: The n^{th} moment about the mean can be defined by 3-11:

$$\mu_n = \sum_{i=1}^{L} (z_i - m)^n \, p(z_i) \quad \text{3-11}$$

The n^{th} moment represents the variance of the intensity distribution for $n = 2$ and skewness of the histogram for $n = 3$.

- Standard deviation: It measures the average contrast in a given image and it is defined as in 3-12:

$$\sigma = \sqrt{\mu_2(z)} \quad \text{3-12}$$

- Smoothness: The relative smoothness of the intensity in a region can be measured by 3-13. The value of *S* is *0* for a region of constant intensity and approaches *1* for regions with large excursions in the values of its intensity levels.

$$S = \frac{\sigma^2}{1+\sigma^2} \qquad 3\text{-}13$$

As it is shown in the next chapters, a combination of the subjective and objective (both unary and binary) measures has been used throughout this thesis to perform quality assessment.

3.3 Dynamic vs. Static Environment (Universe of Discourse)

In addition to the two above discussed types of quality measures we have used another class of quality measures that are application dependent. It means that they are features of objects of interest that should be converted into quality measures. For example for the object of interest of this thesis, face images, such quality measures can be obtained from facial features, e.g. head-pose, and facial components, e.g. eyes and mouth. In the cases that there are more than one quality measures for each image, there should be a way to combine all of these quality measures into one quality score for each image. This will be discussed more in the next chapters.

In addition to the number and types of the quality measures that are involved in the quality assessment, the universe of discourse in which the quality assessment is carried out is of great importance. The universe of discourse determines the method that can be used for converting the quality measures into quality scores. For example suppose that we are assessing the quality of the face images and the assessment is carried out in a static environment that still images are used for quality assessment. If the quality of one of the employed measures is not good enough, the system can ask for a new image from the object. This process can be repeated until the qualities of all the used features are in some pre-defined ranges. These pre-defined ranges are dependent on the application that uses the face images. For example [8] uses a list of facial features and their desired values for the purpose of using in International Civil Aviation Organization applications. Since the acceptable ranges for the quality of the features are known, and we can repeat the process of data acquisition until a good quality for each feature is obtained, the scores assigned to the quality of each feature, in these kinds of environments, can be an absolute score.

In contrast to absolute scoring, relative scoring can be more efficient when face quality assessment systems are working in dynamic environments wherein videos are processed like in surveillance scenarios. In such cases, there is not a possibility for repeating the data acquisition process and therefore, the universe of discourse is limited to the available images in the captured video sequence. Thus, the quality of the images should be compared against each

other rather than against some pre-defined ranges of quality measures of some good quality images. Hence, there is a need for a relative comparison between the available images, i.e. relative scoring. Absolute and relative scoring will be discussed again in the next chapter.

3.4 Summary

This chapter gives a general overview of the quality assessment, its importance, and different types of quality measures. This is necessary to have a good understanding of these measures before going into the details of the proposed systems for face quality assessment in the following chapters. The importance of the environment in which the quality assessment is carried out is also discussed in this chapter.

References

[1] R. de Freitas Zampolo and R. Seara, "*A measure for perceptual image quality assessment*," Proceedings of IEEE International Conference on Image Processing (ICIP), vol. 1, pp. I-433-6, 2003.

[2] Z. Wang and A. C. Bovik, "*Modern Image Quality Assessment*," Morgan & Claypool Publishers, USA, 2006.

[3] R.J. Raghavender, A. Ross, "*Adaptive frame selection for improved face recognition in low-resolution videos*," Proceedings of IEEE International Joint Conference on Neural Networks, pp.1439-1445, 2009.

[4] A. M. Eskicioglu and P.S. Fisher, "*A survey of quality measures for gray scale image compression*," Space and Earth Science Data Compression Workshop, pp. 49-61, 1993.

[5] R.C. Gonzalez, R.E. Woods, and S.L. Eddins, "*Digital Image Processing using MATLAB*," Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2003.

[6] Z. Wang, "*Objective image/video quality measurement – a literature survey*," EE 381K: Multidimensional Digital Signal Processing.

[7] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "*Image quality assessment: From error visibility to structural similarity*," IEEE Transaction on Image Processing, 13(4), pp. 600-612, 2004.

[8] M. Subasic., S. Loncaric, T. Petkovic, H. Bogunvoic, "*Face Image Validation System*," 4th International Symposium on Image and Signal Processing, Croatia, 2005.

CHAPTER 4

FEATURE EXTRACTION

4 Chapter 4: Features Extraction

4.1 Introduction

This chapter is devoted mainly to the second block and the first part of the third block of the proposed system shown in Figure 1-1, i.e. facial feature extraction and normalization of the quality measures, respectively. The extracted facial features will be used as quality measures in the quality assessment. These quality measures will be finally converted into quality scores for each face (It is discussed in the next chapter.). Different numbers of facial features have been used in different face quality assessment systems [1-7, just to name a few]. Griffin [1] uses four features: face pose, expression, background uniformity and lightening. They use these features to choose faces in agreement with ISO standards. Zamani et al. [2] develops a face quality assessment system for measuring the quality of signal during the image acquisition and image restoration from certain noise. They use shadows, hotspots, video artifacts and blurring as quality features. Fronthaler et al. [4] uses the orientation tensor with a set of symmetry descriptors to retrieve the indicators of perceptual quality like noise, lack of structure, blur, etc. Xiufeng et al. [6] develops a framework for face quality standardization. It uses six features: pose symmetry, lightening symmetry, user-camera distance, illumination strength, contrast and sharpness. Subasic et al. [7] defines a face image validation system using 17 facial features. This system checks the usability of still images in travel documents. It assumes the minimum face size of 420x525 pixels. This makes the system inappropriate for working with surveillance videos in which faces are in general smaller than this size. In addition to the resolution problem, most of these systems do not involve features of the facial components in the quality assessment, i.e. they are not robust enough to deal with the different facial expressions which can affect the quality of face images from recognition point of view. Moreover, these systems [1-7] don't use a specific technique for combining the quality scores and usually just use a weighted sum of the features' values.

For the face quality assessment systems proposed in this thesis, several facial features have been employed as face quality measures. This chapter covers the superset of the employed facial features that are used in these systems and their extraction methods. Our different published papers have either used this superset or some of its subsets, as will be discussed in the next chapter. Furthermore, for combining the quality measures and obtaining quality scores for each faces, several methods have been used.

The rest of this chapter is organized as follows: next section describes the superset of the employed facial features and their extraction methods. For extraction of some of the features, more than one method has been explored. Section 3 concludes the chapter.

4.2 Face Quality Measures (Facial Features) and their Extraction

Before going into the details of the facial features it should be mentioned that for extracting some of the features of the detected components, the pixels belong to these components need to be segmented from the irrelevant pixels inside each bounding box. For this purpose, the segmentation algorithm from [8] is used for each detected component: first, the image inside the bounding box, which is produced by the detection algorithm, is divided into objects and background by taking a test threshold. Then the averages of pixels below and above the threshold are computed. Thereafter, the composite average of these two averages is computed by Equation 4-1. Then the threshold is increased, and the process is repeated. Incrementing the threshold stops when it is larger than the composite average.

$$Composite\ Avg. = \frac{background\ avg. + objects\ avg.}{2} \quad 4\text{-}1$$

Segmentation refers to the above process in the rest of this thesis. Figure 4-1 shows an input image, its detected face and facial components and their segmentation results.

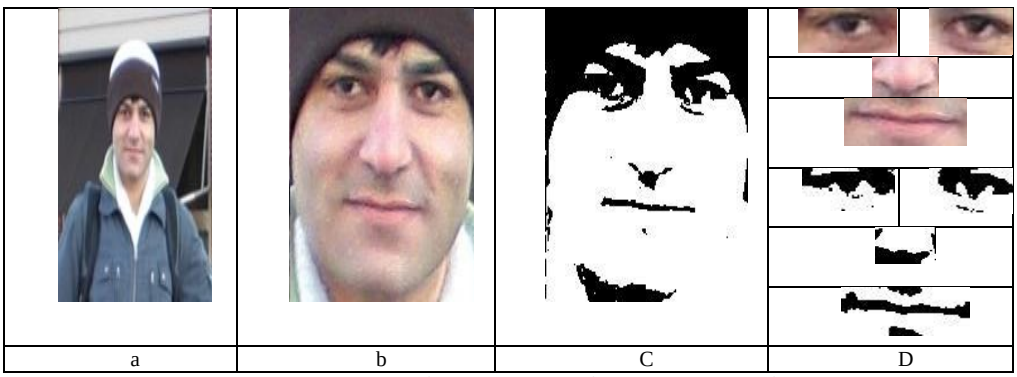


Figure 4-1: a) an input image, b) detected face, c) segmented face, d-top) detected facial components and d-bottom) segmented facial components.

Shape and status of facial components can affect the overall appearance and therefore the quality of face, so it is important to include these features in quality assessment as well as the features of the face. Following is a list of the facial features that have been used as the quality measures throughout this thesis. It has been explained why these features are important for face quality assessment, how they are being extracted and how they are normalized for their universe of discourse that is the entire video sequence.

4.2.1 Face Features

Five features are extracted for the face region: Head-Pose (including two parameters: pan and tilt), Sharpness, Brightness and Resolution.

4.2.1.1 Head-Pose Estimation-Method 1

It is obvious that facial features are more visible in frontal faces than in rotated ones. The more rotated the face the more erroneous the result of the facial analysis systems. Thus, most of the face analysis systems prefer the frontal images to their rotated versions. Usually, people in the video sequences are moving around and looking at different directions, hence their head-poses change. Therefore, it is important to involve head-pose in the quality assessment. The more frontal the face the higher its quality score for this feature. By changes in the head-pose, we mean changes in the tilt and pan of the head.

There are many methods for head-pose estimation in the literature [9-11]. These methods are divided into two classes: local and global methods [10]. Local methods use a set of facial components like eyes, eyebrow and lips to estimate the head-pose. However, estimating these facial components in low-resolution images is difficult. We have used two different methods for head-pose estimation: the first one (method 1) is a global method and the second one (method 2) is a local method that involves the facial features. In the experimental results shown in the next chapter it is explained which of these methods are employed for which systems.

Figure 4-2: Some of the images of one of the training sequences from [12] for the head-pose estimator (method 1).

The first method is a global method. Global methods use the entire face image to estimate the head-pose. These methods typically use a template matching method to find the head-pose. Usually, a nearest neighbor algorithm is applied to find the closest template, and then the pose associated to this template is considered the head-pose. Following [11], we use a set of auto associative memories for estimating the head-pose, one auto associative memory for each pose. Having an input image, the auto associated network with the highest likelihood score is chosen as the head-pose of the input. These memories are trained using the Widrow-Hoff learning rule [11]. This rule provides robustness to partial occlusions. For training the auto associative

memories a public head-pose database [12] has been used. This database consists of 15 sets of images. Each set contains two series of 93 images of the same person at different poses. There are 15 people in the database, wearing glasses or not and having various skin color. The pose, is determined by two angles *(p, t)*, each varying from -90 degrees to +90 degrees. The steps of changes for pan p and tilt t are 15 and 30, respectively. Figure 4-2 shows a training sequence of this database.

Having a video sequence as the input to the system, the values of *p* and *t* are estimated for all the faces of the sequence as explained above. Suppose that for now we work on the pan. The minimum value of *p* in the entire sequence is found and then 4-2 is used in to normalize the value of this feature for all the faces in the sequence:

$$NVF_1 = \frac{P_{min}}{P_i} \qquad 4\text{-}2$$

Input image				
Pan	0	15	30	45
	0+15=15	15+15=30	30+15=45	45+15=60
NVF_1	1	0.5	0.33	0.25

Figure 4-3: Changes in the head-pose (pan) of a face image and correspondingly in the normalized value of the first feature for such a sequence. Since P_{min} is zero the second minimum value of pan in the sequence (i.e. 15) is added to all the values, thereafter, equation 4-2 can be used for normalization.

where NVF_1 is the normalized value of the first feature for the *i*th face image in the sequence, P_{min} is the minimum value of the pan in the sequence, and P_i is the value of pan of the *i*th face image in the sequence. Figure 4-3 shows a video sequence in which the pan of the face of the object and accordingly the value of the normalized value of this feature are changing. If P_{min} is zero the next minimum value is added to all the values and then 4-2 is used for normalization (Figure 4-3). It should be mentioned that this technique of adding the next minimum value in the sequence has been used throughout this thesis wherever the dominator of the normalization equations becomes zero.

By employing the same method, the value of tilt is normalized using 4-3:

$$NVF_2 = \frac{T_{min}}{T_i} \quad 4\text{-}3$$

where NVF_2 is the normalized value of the second feature for the *i*th face image in the sequence, T_{min} is the minimum value of the tilt in the sequence and T_i is the value of tilt of the *i*th face image in the sequence.

4.2.1.2 Head-Pose Estimation-Method 2

The second method for the head-pose estimation that is used in some of the papers that have been produced from this thesis is a local method that involves the facial features. Since this method involves low level features of the face to obtain an exact numerical value for the head-pose, it cannot yield good results in cases of having spectacles or poor quality conditions. On the other hand, in face quality assessment for video sequences obtaining an exact numerical value or a degree of rotation for the head-pose is not needed. Instead, it is needed to choose the least rotated face of a person among the available faces of that person in a given video sequence. Following is the definition of head pan and tilt based on these low level features.

The head pan is defined as the difference between the center of mass of the skin pixels and the center of the bounding box of the face, a coarse estimate which will suffice. For calculating the center of mass, the skin pixels inside the face region should be segmented from the non-skin ones using the segmentation algorithm described in 2.4.1. Then using the following equation the pan value is estimated:

$$P_i = \sqrt{(x_{cm} - x_{bb})^2 + (y_{cm} - y_{bb})^2} \quad 4\text{-}4$$

where P_i is the estimated value of pan of the *i*th image in the sequence, (x_{cm}, y_{cm}) is the center of mass of the skin pixels and (x_{bb}, y_{bb}) is the center of the bounding box of the face. Since the highest score for this feature should be assigned to the least rotated face, again Equation 4-2 is used to normalize the scores of this feature for all the faces in the sequence. Figure 4-4 shows a simple video sequence with three images and the head-pan information that are obtained using the above explained method.

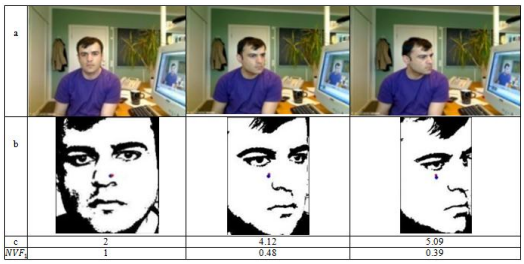


Figure 4-4: Changes in the head pan of images of a given sequence: a) Input sequence, b) detected and segmented faces with center of mass (red) and center of region(blue) marked and c) P_i.

To estimate the head tilt using the low level facial features, the cosine of the angle of the line connecting the center of mass of both eyes is used. These centers are found using the segmentation method explained in section 2.4.1. Having extracted the values of this feature for all the faces in the sequence again Equation 4-3 is used to normalize them. Figure 4-5 shows an example.

Figure 4-5: Changes in the head tilt of a given sequence: a) input sequence, b) detected center of mass of the eyes and c) T_i.

4.2.1.3 Sharpness

Facial details are more visible in sharp images than in blurry images. However, objects in video sequences are usually moving freely in front of the camera, hence having blurry images is often a reality. Thus, it is necessary to involve this feature of the detected faces in the quality assessment. For obtaining the sharpness value of each of the images in the video sequence, we use 4-5:

$$Sh_i = avg(abs(F - lowPass(F))) \qquad 4\text{-}5$$

where F is the ith face image in the video sequence, Sh_i is its sharpness, *lowpass* is a low-pass filter, and *abs* and *avg* are the absolute and average function, respectively [13]. Having

extracted the above value for all the face images of the sequence, 4-6 is used to normalize this feature:

$$NVF_3 = \frac{Sh_i}{Sh_{\max}} \qquad 4\text{-}6$$

where NVF_3 is the normalized value of the third feature for the ith image in the given video sequence, Sh_i is the value of sharpness for this image and Sh_{max} is the maximum value of sharpness in the given video sequence. Figure 4-6 shows a video sequence in which the sharpness and correspondingly the normalized values obtained for this feature from 4-6 are changing.

Input image				
Sharpness	4.14	3.31	3.01	2.96
NVF_3	1	0.79	0.72	0.71

Figure 4-6: Changes in the sharpness of a face image and correspondingly in the normalized value of the third feature of the faces of such a sequence.

4.2.1.4 Brightness

Facial analysis systems have problems when using dark images, because extracting facial features in such images is difficult. In video sequences of moving objects, it is very likely to have changes in the illumination conditions. Therefore, it is important to involve this feature in face quality assessment. The brighter the image, the higher its quality score for this feature. This may risk favoring too bright images that are not good neither for extracting facial features. However, it should be mentioned that, the face detector does not usually detect too bright images.

For obtaining the brightness of the faces in the video sequence, using the segmentation method explained in [8], inside the detected face regions, we first separate the skin pixels from the background and convert them to the *YCbCr* color space and then use the mean of the *Y* component as the brightness of the face. Having extracted these values for all the faces of the sequence, we use 4-7 to normalize them:

$$NVF_4 = \frac{B_i}{B_{\max}}$$

where NVF_4 is the normalized value of the fourth feature for the ith image in the given sequence, B_i is the brightness of the ith image and B_{max} is the maximum value of brightness in the entire video sequence. Figure 4-7 shows a video sequence and the value of the brightness and the normalized value of this feature for each face.

Input image				
Brightness	233	211	183	169
NVF_4	1	0.9	0.78	0.72

Figure 4-7: Changes in the brightness of a face image and correspondingly in the normalized value of the fourth feature of the faces of such a sequence.

4.2.1.5 Resolution

The resolution of an image is simply defined as the multiplication of the height and width of the image. The features of the facial components like the corners of the eyes or the tip of the nose are more easily detectable in high-resolution images than in low-resolution ones. Since objects in the video sequences are moving, their distance to the camera is changing. It causes the head-resolution to change across the video sequence. Hence, it is important to involve this feature in the quality assessment as well. To do so, we find the resolution of all the detected faces in the entire sequence, then give the highest quality score of this feature to the biggest one, and reduce the score as the size of the head is decreasing, as in 4-8:

$$NVF_5 = \frac{R_i}{R_{\max}}$$

where NVF_5 is the normalized value of the fifth feature for the ith image in the given video sequence, R_i is the value of the resolution for this image and R_{max} is the maximum value of the

resolution in the entire sequence. Figure 4-8 shows a video sequence in which the resolution and correspondingly the values obtained for this feature from 4-8 are changing.

The shape and status of facial components can affect the overall appearance and quality of face, so it is important to include these features in quality assessment. Therefore, after extracting the above explained five features for each detected face, we next extract some of the facial components features. These are explained in the following subsections. It should be mentioned that extracting most of these features needs the input image to be of high-resolution. If the input image is not of high-resolution and these features or any other features are not extractable we consider the values of the non-extractable features as zero. Due to the face that our universe of discourse for normalization is the entire input video sequence and since we are using the same methods for all the images of the sequence, we still can be optimistic that having some non-extractable features will not be such a big deal.

Input image				
Resolution	123x169	93x127	62x84	43x58
NVF_5	1	0.56	0.25	0.11

Figure 4-8: Changes in the resolution of a face image and correspondingly in the normalized value of the fifth feature for such a sequence.

4.2.2 Eyes Features

Two features of the eyes are involved in the face quality assessment. Following is the description of these features.

4.2.2.1 Openness of the Eyes

There are plenty of facial analysis systems using the information of iris for identification purposes. Thus, visibility of iris in the face images is important. This visibility depends on the openness of the eyes. Hence, we involve the openness of the eyes in the quality assessment. As a straightforward method for determining the openness of the eye, we have used the aspect ratio of the eye's height to its width. The higher this ratio, the more open the eye, and hence the higher the quality score associated to this feature, correspondingly.

To estimate the region of the iris, we use the Haar-like features first to look for the eyes inside the detected face. Then the previously explained segmentation algorithm from [8] is applied to

the gray-scale version of the eye's region. This algorithm automatically divides the eye's region into the eye and background by taking a test threshold. Then, the averages of pixels below and above the threshold are computed. Thereafter, the composite average of these two averages is computed by 4-1. Then, the threshold is increased, and the process is repeated. Increasing the threshold stops when it is larger than the following composite average. Besides its simplicity, this method is more efficient than methods which apply Hough transform to find the iris itself [14, 15].

Figure 4-9 shows the result of this segmentation. From the segmented image, the height and width of the eye are determined using the highest and the lowest pixels, and the rightmost and the leftmost pixels, respectively. Then 4-9 is used for the normalization of this feature in the given sequence:

$$NVF_{6,7} = \frac{O_i}{O_{\max}} \qquad 4\text{-}9$$

where O_i is the openness of the (right or left) eye of the ith face image in the sequence, O_{max} is the maximum value of this feature (for right or left eye, separately) in the sequence, NVF_6 (left eye) is the normalized value of the sixth feature for the ith face image, and NVF_7 (right eye) is the normalized value of the seventh feature for the same face image.

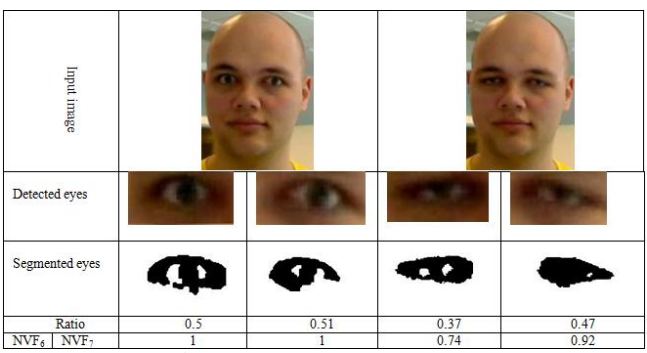


Figure 4-9: Changes in the openness of the eyes and correspondingly in the normalized value of the associated features.

For improving the boundary estimation and noise removal, applying an opening operation to the results of the segmentation can be useful. Figure 4-10 shows such an example for another video sequence with three face images.

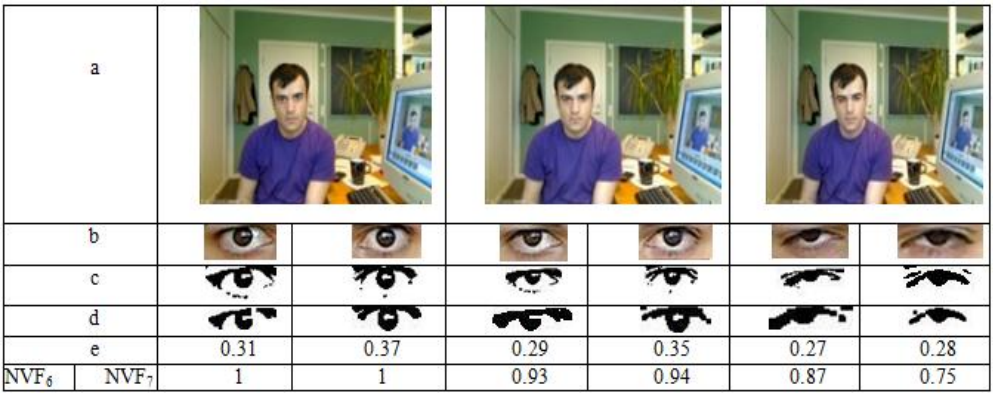


Figure 4-10: Changes in the openness of the eyes: a) input sequence, b) detected eyes, c) segmented eyes, d) opening operation applied to the eyes and e) O_i.

4.2.2.2 Direction of the eyes (Gaze)

The information of the iris is more visible when the person is looking straight into the camera. Furthermore, finding the attention of the people in the face images is always interesting. Thus, there is a need for a feature to assess the gaze of the eye. The more direct the gaze, the higher the value of this feature. For each eye, the direction of the eye (G_i in equation 4-10) is defined as the difference between the eye's center of mass and the center of the bounding box of the eye. The center of mass is found by the same segmentation method describe in the previous section. The Haar features determine the bounding box around the eye.

The importance of this feature is illustrated in Figure 4-11. In these images, the head-poses are the same but the gazes are different. If these two images are the only available images in a sequence, it is obvious that the one that is looking at the camera is preferred to the other.

Having extracted this feature for all the eyes of a person in the given video sequence, 4-10 is used for normalization:

$$NVF_{8,9} = \frac{G_i}{G_{\max}} \qquad 4\text{-}10$$

where G_i is the gaze of the (right or left) eye of the *i*th face image in the sequence, G_{max} is the maximum value of this feature (for right or left eye, separately), NVF_8 (left eye) and NVF_9 (right eye) are the normalized values of the eighth and the ninth feature for the *i*th face image in the sequence, respectively.

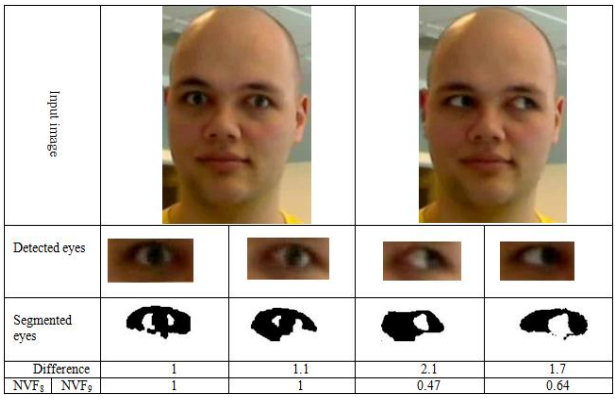


Figure 4-11: Changes in the gaze of the eyes and correspondingly in the normalized value of the associated features.

4.2.3 Nose Feature

The only employed feature for the nose is the aspect ratio of its height to its width. This ratio for a visible frontal or semi frontal nose is between 0.42 and 0.68. Having the nose detected by Harr-like features, if the aspect ratio of the detected nose is far from the above range it will be rejected (i.e. the value of this feature will be considered as zero) otherwise Equation 4-11 is used to normalize this feature in the sequence.

$$NVF_{10} = \frac{NAR_i}{NAR_{\max}} \qquad 4\text{-}11$$

where NVF_{10} is the normalized value of the tenth feature of the *i*th face image in the sequence, NAR_{max} is the maximum value of the nose aspect ratio in the sequence and NAR_i is the value of this feature for current face.

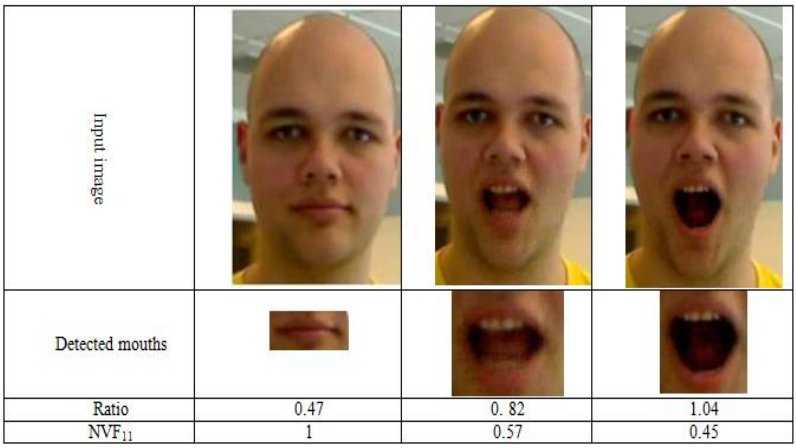
4.2.4 Mouth Feature

Most of the facial analysis systems expect the mouth to be closed in their input images. Changing in the emotional status or expression of the objects in the video sequence can cause changes in the closeness of the mouth. Thus, we involve this feature in the quality assessment and this is the only feature that we consider for the mouth.

To determine the closeness of the mouth, we first detect it using the Haar-like features and then consider the ratio of the mouth's height to its width as the value of this feature (Figure 4-12). The closer the mouth, the smaller this ratio. Thus, having extracted this feature for the mouths of all the detected faces, 4-12 is used for normalization:

$$NVF_{11} = \frac{C_i}{C_{\max}} \qquad 4\text{-}12$$

where NVF_{11} is the normalized value of the 11th feature of the ith face image, C_i is the value of the closeness of the mouth for this image, and C_{min} is the minimum value of this feature for all the mouths detected from the given video sequence.



Ratio	0.47	0. 82	1.04
NVF_{11}	1	0.57	0.45

Figure 4-12: Changes in the openness of the mouth and correspondingly in the normalized value of its associated feature.

4.3 Summary

This chapter defines the superset of facial features that are used as the quality measures for face quality assessment in this thesis. It is explained why these features are important for face quality assessment and how they are being extracted. Furthermore, the normalization method for each of these features is given. Different papers produced from this thesis either have used this superset or they have employed a subset of this superset.

References

[1] P. Griffin, "*Understanding the Face Image Format Standards*," National Institute of Standards and Technology Workshop, American National Standards Institute, Gaithersburg, Maryland, USA, 2005

[2] A.N. Zamani, M.K. Awang, N. Omar, and S.A. Nazeer, "*Image Quality Assessments and Restoration for Face Detection and Recognition System Images*," IEEE 2nd Asia International Conference on Modeling & Simulation, Kuala Lumpur, Malaysia, 2008.
[3] Q. Xiong and C. Jaynes, "*Mugshot Database Acquisition in Video Surveillance Networks Using Incremental Auto-Clustering Quality Measures*," IEEE Conference on Advanced Video and Signal-based Surveillance, Miami, USA, 2003.
[4] H. Fronthaler, K. Kollreider and J. Bigun, "*Automatic Image Quality Assessment with Application in Biometrics*," IEEE Conference on Computer Vision and Pattern Recognition Workshop, New York, USA, 2006.
[5] Y. Rodriguez, F. Cardinaux, S. Bengio and J. Mariethoz, "*Estimating the quality of face localization for face verification*," IEEE International Conference on Image Processing, Singapore, 2004.
[6] G. Xiufeng, Z. Stan, L. Rong, and P. Zhang, "*Standardization of Face Image Sample Quality*," International Conference on Advances in Biometrics, Seoul, Korea, 2007.
[7] M. Subasic, S. Loncaric, T. Petkovic and H. Bogunvoic, "*Face Image Validation System*," 4th International Symposium on Image and Signal Processing and Analysis, Zagreb, Croatia, 2005.
[8] M.I. Rajab, M.S. Woolfson and S.P. Morgan, "*Application of Region-Based Segmentation and Neural Network Edge Detection to Skin Lesions*," Journal of Computerized Medical Imaging and Graphics, vol. 28, pp. 61-68, 2004.
[9] K. Nasrollahi and T.B. Moeslund, "*Face Quality Assessment System in Video Sequences*," 1st European Workshop on Biometrics and Identity Management, Roskilde, Denmark, 2008.
[10] E. Murphy-Chutorian and M. Trivedi, "*Head Pose Estimation in Computer Vision: A Survey*," IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 31, no. 4, pp. 607-626, 2008.
[11] N. Gourier, J. Maisonnasse, D. Hall, and J.L. Crowley, "*Head Pose Estimation on Low Resolution Images*," Stiefelhagen, R., Garofolo, J.S. (eds.) CLEAR 2006. LNCS, 4122, 270-280, Springer, 2007.
[12] N. Gourier and J. Letessier, "The Pointing 04 Data Sets," International Conferemce on Pattern Recognition, Visual Observation of Deictic Gestures, UK, 2004.
[13] F. Weber, "*Some Quality Measures for Face Images and Their Relationship to Recognition Performance*," Biometric Quality Workshop, National Institute of Standards and Technology, USA, 2006.
[14] Q. C. Tian, Q. Pan, Y. M. Cheng and Q. X. Gao, "*Fast Algorithm and Application of Hough Transform in Iris Segmentation*," the IEEE 3rd International Conference on Machine Learning and Cybernetics, Shanghai, 2004
[15] J. Huang , Y. Wang, T. Tan and J. Cui, "*A New Iris Segmentation Method for Recognition*," the IEEE 17th International Conference on Pattern Recognition, Huang, 2004

CHAPTER 5

SCORING AND FACE-LOG GENERATION

5 Chapter 5: Scoring and Face-Log Generation

5.1 Introduction

This chapter is devoted to the second part of the third block and the first two parts of the fourth block of the proposed system shown in Figure 1-1. Having extracted the facial features explained in the previous chapter, now it is the time for converting them into a quality score for each face of the input video sequence. This process is denoted *scoring* throughout this thesis. Most of the face quality assessment systems in the literature [1-7, to name a few] use a simple weighting system for scoring. Unfortunately, most of these systems do not explain how they obtain the weights for the quality measures. During the completing of this thesis we found out that the effectiveness of scoring highly depends on the application that is going to use the result of the face quality assessment. If the results of the face quality assessment is going to be used for example for video indexing a simple weighting system may suffice. However, if the results of the quality assessment are going to be considered as a complete and concise [8] summarization of the input video sequence, more advanced techniques are required. Furthermore, if the results of the face quality assessment system are going to be used in applications like super-resolution some more post processing and some more quality measures, like binary objective quality measures are required. These will be explained more in this chapter and in chapter 7.

The rest of this chapter is organized as follows: the next section is devoted to the definition of face-logs. Section 3 discusses some of the applications of the face-logs and explains the experimental results of some of the related published papers of this thesis. These systems use different techniques for scoring. Finally, section 4 concludes the chapter.

5.2 Face-Log

Face-logs are considered as a concise and/or complete representation of a video sequence. This term was first introduced by [8] and then developed in [9-11]. In Zhang and Wang [10] to solve the incapability for performance prediction and remove the requirement for scale normalization of existing methods, three face quality measures are used. Bilateral asymmetry of SIFT feature point is utilized to extract scale insensitive feature points on face images, and three asymmetry based quality measures are calculated by applying different constraints. Bimbo et al. [11] propose an intrusion logger. An intrusion logger is a video surveillance application designed to detect intrusion events and document them by storing time stamped images in a log. Face-loggers, in particular, are focused in grabbing imagery of intruders face, and are typically used to provide inputs to a face recognition system. Commonly, intrusion loggers are supposed to signal an alarm condition as the intrusion occurs, and are therefore required to run in real time and to work continuously for long time periods. The face-logging solution proposed in [11] is capable of detecting and tracking several targets in real time, grabbing face images and

evaluating their quality in order to store only the best for each detected target. However, as it is shown in the following sections keeping only the best face image is not always sufficient.

5.3 Face-logs for different purposes

The content of face-logs depends on the application. For example if the face-logs are used for indexing video sequences, they may only contain the best face image of the sequence. If they are used for summarizing a video sequence, they should be complete. It means that they should contain the best side-view images (if any) as well as the best frontal image. If face-logs are used for super-resolution, as in this paper, in addition to the best frontal (and side-views) face(s) they need to have more images. These images should be closely similar to the best image. These are explained in the following subsections.

5.3.1 Best Face Image(s) (BFI)

Best face image can be a good representative for a given video sequence and can be used for video indexing or for face recognition instead of applying the face recognition algorithm to all the images of the video sequence. By changing the different factors involved in face quality assessment, different systems have been proposed for finding the best face images of a given video sequence. Following is the descriptions of these systems.

5.3.1.1 BFI-System 1

Our second proposed system (BFI-System 1) [12] uses four facial features as quality measures for face quality assessment. These features are: head-pan, sharpness, brightness and resolution. These features are extracted using the methods explained in 4.2.1.2, 4.2.1.3, 4.2.1.4, and 4.2.1.5, respectively. For scoring, this system uses a weighting system. Hence, for each detected face image the extracted features are combined into a quality score using the following equation:

$$QS_i = \frac{\sum_{j=1}^{4} w_j NVF_j}{\sum_{j=1}^{4} w_i}$$

where QS_i is the quality score for the ith face image in the sequence, NVF_j are the normalized values for the extracted facial features, and w_j are the weights of the features. The experimentally obtained values of weights are shown in Table 5-1.

Table 5-1: The values of the weights of the quality measures for BFI-system 1

Weight	w_1	w_2	w_3	w_4
Value	1.8	0.9	0.6	0.8

The images are sorted based on their combined scores obtained from 5-1. Then, depending on the application, one or more images with the greatest values in *QS* are considered as the highest quality image(s) in the given sequence.

5.3.1.1.1 BFI-System 1: Experimental Results

We have used both still images and movie samples to evaluate our system. Following is a brief description of these databases and then the results of the system using them are presented.

FRI CVL (DB3): This database [13] contains 1008 still images. This database consists of sequences of 114 individuals. Each sequence has 7 images with different head rotation (Figure 5-1). Since the images in this database do not have wide variations in sharpness, brightness and size we have used them mainly for assessing the first feature. If the other features of the face are not changing, the least rotated face can give us the best face in terms of the visibility of the facial features. These color images have been taken by a Sony Digital Mavica and the size of the images is 640x480 pixels.

Hermes Database (DB4): In order to assess the other features as well as the first feature, we have used a video dataset that is prepared for Hermes project [14]. This dataset contains 48 sequences (6 videos for each of the 8 participants) where these people walk towards a camera while looking from side to side. This provides good examples for assessing all the features together.

In order to validate the results of the proposed system, we have asked an expert to annotate the images in each sequence in our datasets according to their visual features and the visibility of the face and sort them manually based on his perception of the quality. Table 5-2 shows the results of this comparison using these two databases, in which, the correct matching means the matching between the human perception and system results for the best images in each sequence. While the quality of the images is not too poor and the faces size is not too small the ranking order of the selected images by the proposed system is similar to the order of the selected images by the human. By the way, even for the poor quality images, although it is possible that the images be sorted in different way by the system and the expert, in most of the cases we can find the best chosen image by the human inside the first four chosen images by the system.

Table 5-2: Experimental results for BFI-System 1

Database	Number of sequences	Number of faces in sequences	Face detection rate	Correct matching
DB3	114	7	94.3%	92.1%
DB4	48	avg. 15	90.5%	87.1%

Figure 5-1 shows the results of the quality-based ranking by the proposed system and the human for an examples from the FRI CVL dataset. In general the human and system rankings are in agreement. Slight differences like those seen in the figure occur when the images in the database are very similar e.g., like the three in the center.

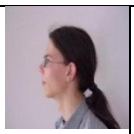
a							
b	5	4	1	2	3	4	5
c	5	3	1	1	2	4	6

Figure 5-1: An example from the FRI CVL database (DB3) and the quality-based rankings: a) the input sequence of still images, b) Human ranking and c) System ranking.

Figure 5-2 shows a sequence of images from the Hermes dataset and the results of sorting their faces based on the quality both by a human and the proposed system. It is obvious from these images that the selected faces by the system match to the selected faces by the human for the first five images.

							
							
8	7	6	5	4	3	1	2
8	6	7	5	4	3	1	2

Figure 5-2: Quality-based rankings for a sequence from Hermes dataset (DB4): a) input video sequence, b) Extracted Face, c) Human ranking, and d) System ranking.

Figure 5-3 shows another example from the Hermes dataset. In this sequence the size of the head is not changing widely. However, since the person turns around his head while he is walking, the other features are changing. It can be seen that in this case the most important feature is head rotation and the proposed system ranking has an acceptable agreement with the human ranking.

a							
b							
c	5	3	2	1	6	4	7
d	4	2	3	1	6	5	7

Figure 5-3: Quality-based rankings in the presence of head rotation for another video sequence from Hermes dataset (DB4): a) input video sequence, b) Extracted Face, c) Human ranking, and d) System ranking.

Figure 5-4 shows another example from the Hermes dataset in which the quality of the images are very poor and the walking person has spectacles. In this figure the details of the normalized features and also the combined scores are shown.

a							
b							
c	7	6	5	4	3	2	1
NVF_1	0.25	0.14	0.17	0.27	0.71	0.37	1
NVF_2	0.73	0.75	0.75	0.86	1	0.89	0.85
NVF_3	0.95	0.94	0.93	0.95	0.97	0.98	1
NVF_4	0.2	0.28	0.37	0.51	0.64	0.96	1
QS_i	0.49	0.48	0.51	0.61	0.81	0.76	0.95
d	6	7	5	4	2	3	1

Figure 5-4: A poor quality sequence of images from Hermes database (DB4) and the details of the locally

scoring technique: a) the input video sequence, b) the detected faces, c) human ranking, NVF_i the jth {j=1..4} face quality measures, QS_i the quality score of the ith face image in the video sequence and d) the ground truth.

As seen in the above figures (Figure 5-1 to Figure 5-4), the quality-based rankings by the proposed system and the human are very close. A few incorrect ordering were observed which are due to the fact that BFI-System 1 cannot detect the exact direction of the face, as well as the facial expressions. When the images in the sequence are very similar and the face image are too small then the possibility of miss ranking by the system increases. However, in general very good results are obtained.

5.3.1.2 BFI-System 2

Our third proposed system (BFI-System 2) [15] that has been developed in this thesis for finding the best face image(s) of a given video sequence uses the entire superset of the facial features explained in previous chapter. This system uses the second method for head-pose estimation that is described in 4.2.1.2. The system works with high-resolution images better than low-resolution ones, because it involves the features of the facial components as well as the features of the face. However, since the universe of discourse is the entire video sequence and since all the images of the sequence are treated in the same way, the system does not fail if some of the facial features are not extractable. In cases like that we just simply consider zero for the values of the non-extractable features.

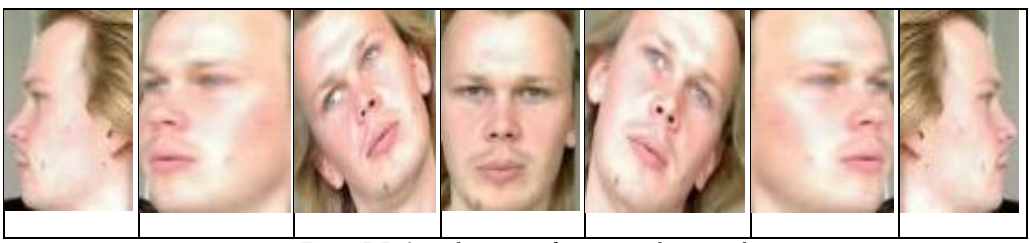


Figure 5-5: A used sequence for training the network.

Having the facial features extracted and normalized, we have used a neural network (multi-layer perceptron (MLP)) for *scoring*. This neural network gets the quality measures as input and produces a quality score for each face image. Therefore, it has 11 inputs and one output. For training this neural network 140 face images of 20 persons from a local database are used. Figure 5-5 shows one of these training sequences in which some of the facial features are changing. The other training sequences provide enough examples for the network to learn all the involved features. While training the network, for each of these images, the following equation is used to produce a desired value for each face:

$$QS_i = \frac{\sum_{j=1}^{11} w_j NVF_j}{\sum_{j=1}^{11} w_j} \qquad 5\text{-}2$$

where, QS_i is the quality score of the *i*th face image in the video sequence, NVF_j is the normalized value of the *j*th feature, where *j= {1..11}* and w_j is the weight associated to the *j*th feature.

For determining the weights, all of them are first initialized to two and then some of them are decreased by steps of 0.1 to 1. In each step the network is trained and its learning rate is monitored. The best learning rate of the network obtains by the weights shown in Table 5-3.

Table 5-3: The weights obtained for the quality measures in BFI-System 2

j	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*
w_j	2	1.7	1.4	1	1.6	1.4	1.4	1.4	1.4	1	1.5

The network is trained by the features from the 140 training images as the input and the value for these features from Equation 5-2 as the desired value.

5.3.1.2.1 BFI-System 2: Experimental Results

Having a video sequence as the input the above mentioned 11 features are extracted and normalized for each face in the sequence and then fed to the MLP to produce a quality score for each image. Based on these quality scores, the system is giving a ranking number to each face. These ranking numbers show the priority of images for participating in face-logs generation. These numbers increase as the quality scores decrease. The face-log for each person is generated by including his/her best images according to the above ranking numbers.

In order to validate the proposed system, four databases are used where three of them are public available. The same as the previous system, these databases are annotated by human experts based on the quality perception of face images and the visibility of the facial components. These annotated data are referred to as ground truth and the results of the system are compared against them. In the following these databases are first described and then the results of the system using them are showed. Finally the results of the proposed system are compared against two of the most similar systems in the literature.

Table 5-4 shows the time needed by each block of the system for the biggest (960x720 pixels) and smallest (92x112 pixels) input images. As it can be seen from this table, the proposed system is working in real time.

Table 5-4: The time needed by different parts of the system in ms. TBI and TSI are the time needed for processing of the biggest and the smallest input images.

	TBI(ms)	TSI(ms)
Detection	53.1	10.7
Feature Extraction	15.7	8.6
Scoring	2.4	2.4
Total	71.2	20.7

FRI CVL Database (DB3): For description of this database please refer to the previous section. Figure 5-6 shows one sequence of this database and the resulted rankings of the proposed system which are in general in agreement with the ground truth, except for the two last ones which are not detectable by the employed face detector.

a							
b	-						-
c	-	5	3	2	1	4	-
d	-	5	3	2	1	4	-

Figure 5-6: The results of the system for DB3: a) input images, b) detected faces, c) ground truth and d) system results.

The first row of Table 5-5 shows the results of the system for this database. Due to the bigger sizes of the images, the results of this database are better than the other databases.

Table 5-5: Face-Logs (containing the m-best images) matching rates between the results of the system and the ground truth in each database.

	1-best	2-best	3-best	4-best
DB3	100%	93.2%	91.4%	85.4%
DB5	100%	90.1%	89.5%	82.1%
DB6	100%	87.4%	85.2%	80.8%
DB7	98.1%	83.2%	79.1%	71.3%
Average	99.5%	88.4%	86.3%	79.9%

AT&T Database (DB5)*:* This database [16] contains 40 distinct subjects. The size of each image is 92x112 pixels. All the images have been taken against a dark homogeneous background with the subjects in an upright, frontal position with tolerance for some side movement. Changes in the facial expression and the status of the facial components make this database suitable for face quality assessment. Figure 5-7 shows one sequence of this database and the ranking results of the proposed system which is in general in agreement with the ground truth.

Figure 5-7: Results of the system for DB5: a) input images, b) Ground Truth and c) System Results

The second row of Table 5-5 shows the results of the system for this database. In this table the matching between the results of the system and the ground truth, for the face-logs containing the *m*-best images (*m*=1, 2, 3, 4) is shown.

Face96 Database (DB6): This database [17] contains 152 sequences each consisting of 20 images. The subjects in this database are walking towards a fixed camera. There is 0.5 seconds between the successive frames. Background is complex and head scale is changing while the head tilt, turn and slant have minor variations. Size of the faces in these images is changing from 60x65 to 80x95.

Figure 5-8 shows every other image of a sequence of this database and the results of the proposed system vs. ground truth.

Figure 5-8: Results of the system for DB6: a) input images, b) detected faces, c) ground truth and d) system results.

The third row of Table 5-5 shows the results of face-log matching between the proposed system and ground truth for this database.

Local Database (DB7): The images in the first database are changing widely in head rotation but the distance between the camera and the subjects is fixed. The face images in the second database have good variations in facial components like opening the mouth, closing the eyes and gaze, but the head rotation, head scale and lightening are not changing that much. In the third database even though subjects are moving towards the camera and the background is complex, the facial expression is not changing.

In order to involve all the features of interest in the assessment process another dataset (DB7) is used. Figure 5-9 shows an example from this sequence. The sequences in this database are more realistic compared to the other databases. The 20 subjects participating in this database are being asked to talk, change their gaze and head rotations while moving towards a camera. For each of these 20 people two sequences are extracted each containing almost 50 frames. The last row of Table 5-5 shows the results of face-logs matching using this dataset.

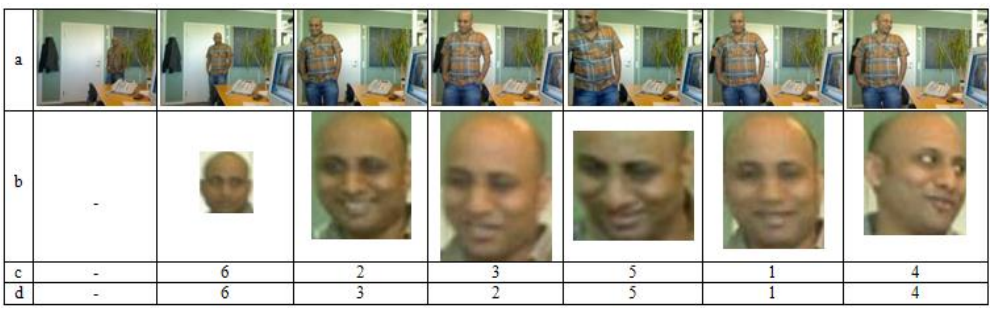


Figure 5-9: The results of the system for DB7: a) input images, b) detected faces, c) ground truth and d) system results.

The systems proposed in [7] and [8] are among the best face quality assessment systems in the literature. [7] uses 17 features to assess the quality of face images in an static environment to decide if a specific image is suitable for use in travel documents or not. [8] uses five features to assess the quality of faces in a video sequence. The above systems are simulated by changing the proposed system and including only the features that they use. Then these simulated systems are applied to DB7 and the results of the matching rates between the systems and the ground truth for constructing face-logs containing first and second best images are shown in Table 5-6. For simulating system [7] we have included all the features except the features relating to the background and the color information. Because there is no background in the detected faces and the proposed system is working with gray images as well as color ones. We

have included all the features that are involved in system [8] except feature relating to the human skin.

Table 5-6 shows that the proposed system is performing much better than these two systems. The reason for that is the content of database DB7, in which the facial components like eyes and mouth, head yaw and tilt are changing as well as the other facial features. Since these two systems, especially system [8], do not involve these features completely, they have difficulties in finding the best images.

Table 5-6: Comparing the proposed system vs. state of the art systems

Database	1-best	2-best
System [7]	85.5%	78.5%
System [8]	79.4%	70.1%
Proposed System	98.1%	83.2%

5.3.2 Complete Face-Log (CFL)

The two systems proposed in the previous section are both good for face quality assessment and both can extract the best face image of a given video sequence, perfectly. It means they can produce face-logs containing only the best face-log. However, if we want to use them for producing face-logs containing more images than the best face image, like *m*-best face image (*m>1*), they simply choose the *m* first image with the highest quality scores. It means that there would be a high degree of redundancy among the members of the log. This is due to the fact that subsequent frames are similar to each other and they get similar quality scores. Therefore, the above two systems cannot produce a complete face-log for a video sequence. A complete face-log is a summarization of a video sequence in which the redundancy of information among the members of the log is minimal.

Following is the description of two new face quality assessment systems. In addition to the fact that these two systems are able to find the best face image of a video sequence, they can produce complete face-logs for their inputs.

5.3.2.1 CFL-System 1

Similar to BFI-System 1, our fourth proposed system (CFL-System 1) [9] uses four facial features as quality measures for face quality assessment. These features are: head-pan, sharpness, brightness and resolution. These features are extracted using the methods explained in 4.2.1.2, 4.2.1.3, 4.2.1.4, and 4.2.1.5, respectively. For scoring, however, this system does not employ the same weighting system as BFI-System 1. Instead, it uses a fuzzy inference engine. Having the quality measures extracted and normalized, they are fed to the fuzzy engine. Then, the engine produces a quality score at its output. The reason for fuzzy description of the scores

is to have similar quality scores for similar images. It helps us to obtain a quality score graph that is flat in the different temporal periods, which makes it much easier to find the local maximums in the graph.

5.3.2.1.1 Fuzzy Inference Engine

The used fuzzy inference engine is a Mamdani model [18, 19] with four inputs associated to the above quality measures. For each input we have defined two membership functions that define the strength and weakness of the features (Figure 5-10). For the first input which is related to the pose estimation we have used Gaussian membership functions. It means that the more frontal the face the higher the score associated to this feature. For the next two features, sharpness and brightness, we have used Gaussian bell membership functions. It means that for both membership functions of these two features by improving (or degradation) their associated features their fuzzy scores are increasing up (or decreasing) to a specific limit only, and beyond that limit the feature is good enough (or quite weak). For example for the second feature, sharpness, if the value of the relative score is increasing from 0.4 to 0.65 its fuzzy score increases accordingly. But changing the relative score from 0.65 to 1 has no important effect on the associated fuzzy score. Smaller values of the relative score of this feature are handled by the other membership function of this input. If the value of this relative score is decreasing from 0.4 to 0.25, its fuzzy score decreases accordingly. But if the relative score is less than 0.25 its fuzzy score is quite weak anyway. For the last input, resolution, we have used one Gaussian and one Gaussian bell membership function.

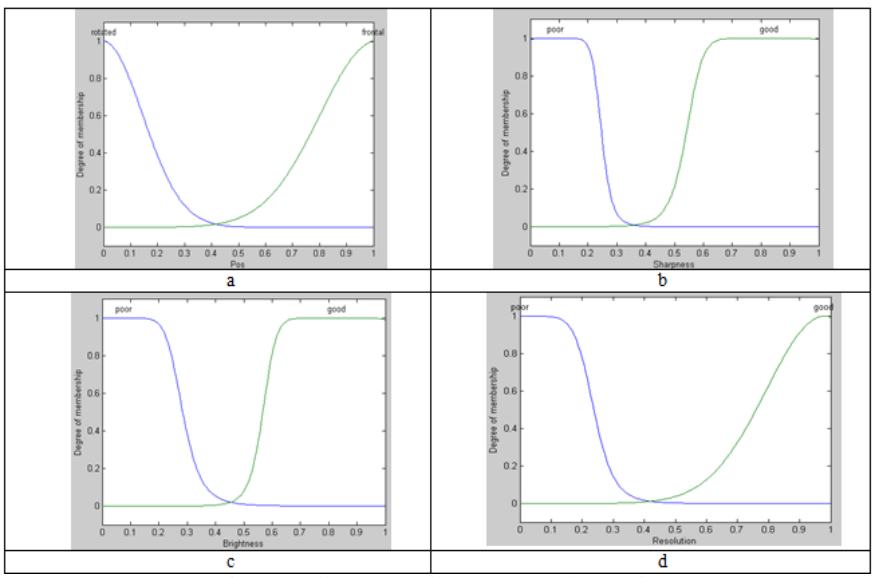


Figure 5-10: The membership functions of the inputs of the employed fuzzy inference engine: a) head-pose, b) sharpness, c) Brightness, and d) Resolution.

This fuzzy inference engine is using the rules shown in Table 5-7. The weights of all the rules are equal and considered as one. The used aggregation method for the rules is Maximum Value [18] and the defuzzification method is the bisector of the area [18]. The membership function associated to the single output of this engine is shown in Figure 5-11.

Table 5-7: The rules used in the fuzzy inference engine.

Rule	*Pos*	*Sharpness*	*Brightness*	*Resolution*	*Quality*
1	rotated	Poor	poor	poor	poor
2	rotated	Poor	poor	good	poor
3	rotated	Poor	good	poor	poor
4	rotated	Good	poor	poor	poor
5	frontal	Poor	poor	poor	average
6	frontal	Good	poor	poor	poor
7	frontal	Poor	good	poor	poor
8	rotated	Good	good	poor	poor
9	frontal	poor	poor	good	poor
10	rotated	good	poor	good	poor
11	rotated	poor	good	good	poor
12	frontal	good	good	poor	average
13	frontal	good	Poor	good	average
14	frontal	poor	Good	good	average
15	rotated	good	Good	good	average
16	frontal	good	Good	good	good

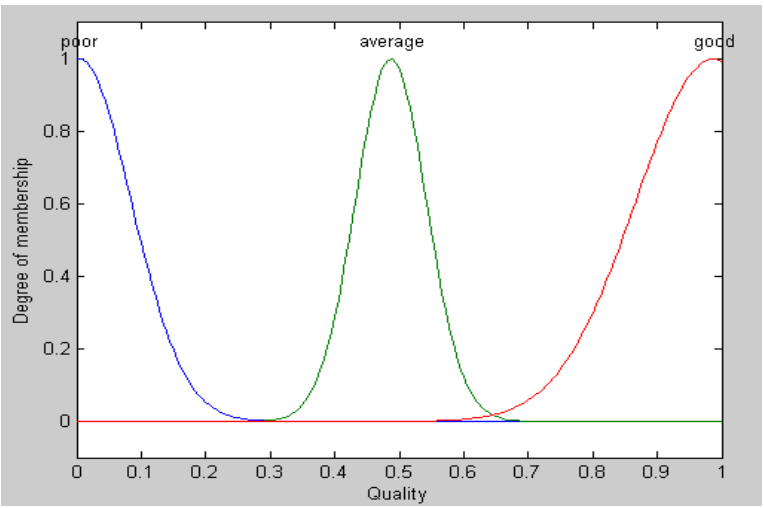


Figure 5-11: The membership functions for the single output of the FIE.

Figure 5-12, shows how the output of the engine changes according to the changes of the individual features and also some of the features simultaneously.

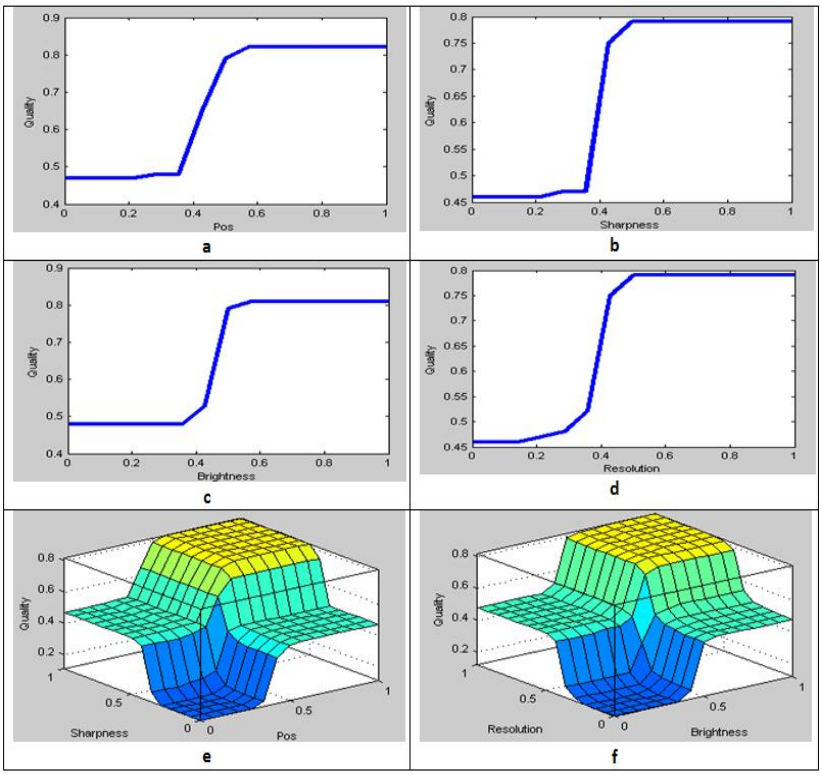


Figure 5-12: Changing of the output of the FIS with respect to the changing of the input features. a-d: Quality vs. Individual features, e-f: Quality vs. two of the features.

In constructing a face-log we like to reduce the redundancy and at the same time build the log in such a way that it contains useful information about the face in the sequence. The output of the above fuzzy inference engine for each found and tracked face in the given sequence will be stored separately. Figure 5-13 shows a simple example in which the output of the fuzzy inference engine is stored for the face regions of one observed person in a sequence with 50 frames. An obvious way for choosing the images to build the log is to choose the *m* images with the highest value in quality score. Let say that frame *x* contains a good face, it is highly possible that frame *x-2, x-1, x+1, x+2* contain a face similar to frame *x* and the quality scores for all of them therefore would be similar. It means that if we just simply choose the *m* frames with the highest score in quality it is likely that the images in the log are very similar to each other and so there is a high degree of redundancy. Instead, *we find the m-local maximum* in the graph and add the images associated with these values to the face-log. This is discussed more in the experimental results.

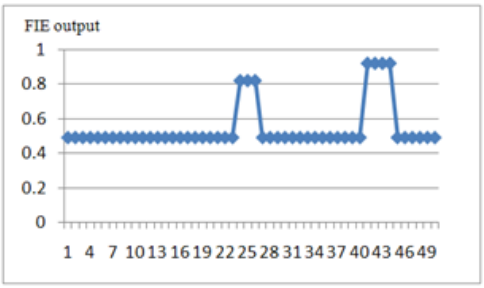


Figure 5-13: The output of the FIE for a given video sequence with 50 frames.

5.3.2.1.2 CFL-System 1: Experimental Results

Similar to the previous described systems, in this section, we first give an overview of the used databases and then the experimental results are given. For testing this system we have used four databases. The first two databases are FRI CVL (DB3) and Hermes (DB4) which have been explained in 5.3.1.1.1. The other two databases are:

Oxford Database (DB8): This database is provided by the Active Vision Group at Oxford University especially for our work. This dataset consists of two long video sequences containing two persons in a very complicated background and hard situations for detecting and tracking the faces. The persons are detected and tracked using [20]. A pan-tilt-zoom camera then zooms in on the upper part of the bounding box. Within this bounding box the Viola-Jones face detector [21] is applied. When a face is detected a general level-set tracker [22] takes over. It tracks the head even during rotation and scale changes. The output is a sequence of stabilized close-up images of the head. This dataset provides a unique (and realistic) challenge due to the involvement of the pan-tilt-zoom camera.

Local Database (DB9): For preparing this database we have used a webcam as a surveillance camera. The 10 subjects participating in this dataset have been asked to do random head and body movements in front of the camera in poor lightning conditions, for 15 seconds each. The random movements provide an opportunity to challenge all the different aspects of the system and validate its reliability.

It is obvious that a system which is trying to build a face-log of *m*-best images must be able to find the best image beforehand. We have therefore provided the experimental results in two parts. In the first part we evaluate our system on short sequences to find the best image. In the second part we show that our system can build complete face-logs containing the *m*-best images in longer sequences.

The most similar work in the literatures is the system described in [8] which builds a face-log containing the best image. In the two following parts we compare our proposed system (CFL-

System 1) with the system described in [8]. In order to validate the results of these two systems we compare them with hand annotated data, i.e. ground truth.

Table 5-8 illustrates the comparison of the results of these two systems with ground truth. In this table the correct matching means matching rate of the labels between the ground truth and systems' results, for choosing the best image. While the quality of the images is not too poor and the face size is not too small the systems' image selection orders are similar to ground truth. For the poor quality images, it happens that the images are sorted in different ways, which results in the drop in the matching rate in the table, but in most of the cases we can find the best image among the first four chosen images by the systems. In general, the quality-based rankings by the systems are close to ground truth. Some incorrect orderings are sometimes observed which are due to the facts that the systems cannot detect the exact direction of the face and the facial expressions. When the images in the sequence are very similar and the face image are too small then the possibility of miss ranking by the systems increases. However, in general good results are obtained (The images for this part of the experiment are the same as the images presented in section 5.3.1.1, therefore, they are deleted to avoid repetition of similar information).

Table 5-8: Comparing the results of CFL-System 1 and System [8] vs. the ground truth.

Database	*Number of sequences*	*Number of faces in sequences*	*Face detection rate*	*CFL-System 1 Correct matching*	*System [8] Correct matching*
DB3	114	7	94.3%	93.4%	92.1%
DB4	48	avg. 15	90.5%	88.5%	87.1%
DB8	45	avg. 50	90.3%	87.9%	86.4%
DB9	100	avg. 50	89.8%	88.3%	87%

Both of these two systems, CFL-System 1 and System [8], have the same performance when they are trying to find the best image, even though we use relatively simple features compared to system [8]. However, if we use these systems for building complete face-logs the results of the two systems are different.

For the purpose of constructing face-logs containing *m* best images, system [8] simply chooses the *m* images with the highest quality score. Although the images selected in this way by system [8] have the highest quality they are simply sequential frames and completely similar to each other. Thus, there is a high degree of redundancy in the constructed face-log, see Figure 5-14. Instead, in our approach we find the *m* highest local maximums in the quality score graph and add the corresponding faces to the face-log. As illustrated in Figure 5-14 and Figure 5-15, using the quality score graph obtained by our system, in addition to having the best image from system [8], the second and third best images which are associated to the other local maximums in the graph, are selected to be added to the face-log. Hence, while reducing the redundancy, the face-log is also complete. The reason for that is the fact that fuzzy quality

scores for the similar images are the same and the local maximums can be found easily in our system. Whereas finding local maximums for system [8] would be difficult, because even a small change in the features can yield a new point in the quality score graph.

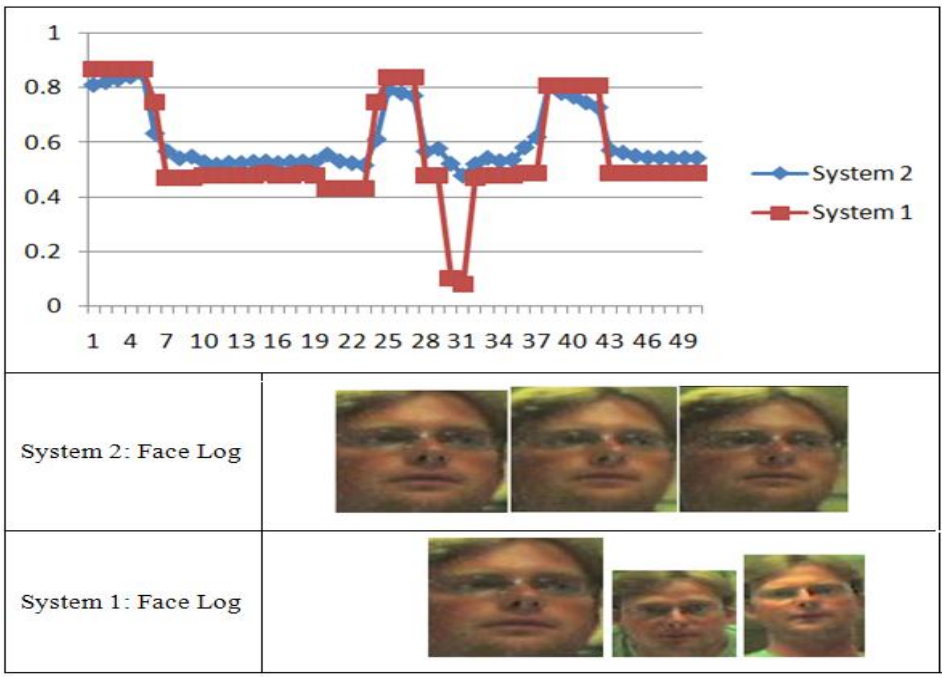


Figure 5-14: From top to down: Quality score graphs for the two systems for a video sequence of almost 50 frames and the (*m=3*)-best chosen images by the two systems for building the face-logs.

We have tested the purposed idea over all our three databases of video sequences (DB4, DB8, and DB9). The results show that this idea makes face-logs more complete and concise than system [8]. Figure 5-14, shows an example from DB8 where the face-log constructed by our system is more complete than the log provided by system [8], while ours does not have any redundancy of information. Suppose you want to use these face-logs to construct a 3D model of the face or use it in an authentication system, then the importance of having the best images from different views or temporal situations will be clearer. Figure 5-15, shows two examples from DB9 (one with 50 frames and another with 45 frames). As can be seen from the figure, the best image selected by two systems is the same, but for constructing the face-log, system [8] simply adds sequential images after the best image, which results in high degree of redundancy. And only after considering all the similar best images (here *3*-best images) this system finds alternative best images located at the other temporal situations. After finding the best image from the first temporal situation, our system, immediately, goes to the other temporal situations and considers the last images added by system [8], as its second choice. This results in less redundancy and in general much more concise face-logs.

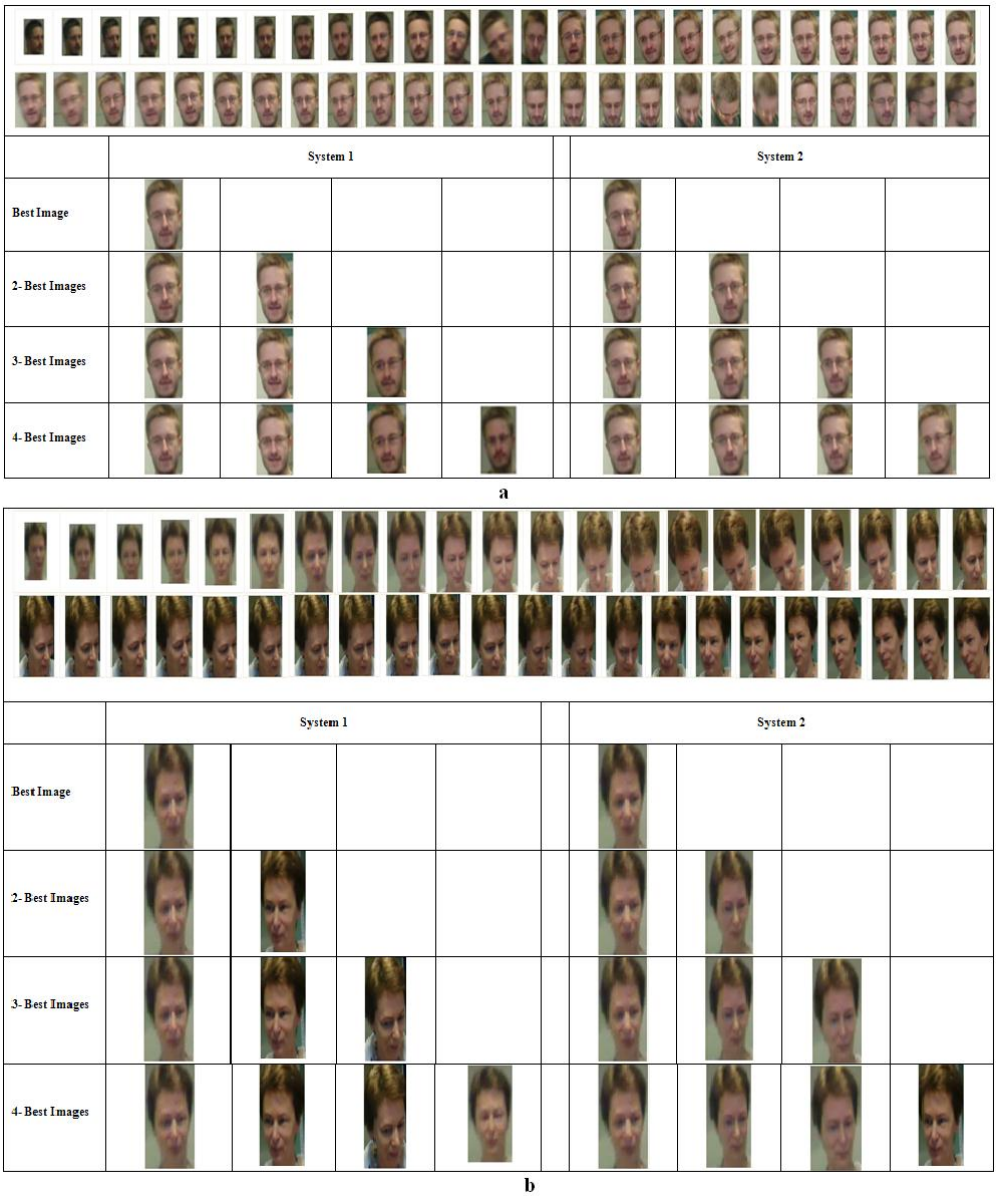


Figure 5-15: Two video sequences from DB9 with a) 50 frames and b) 45 frames and construction of the face-logs with different number of best images by both systems.

5.3.2.2 CFL-System 2

The first system for constructing complete face-logs uses the fuzzy descriptions of the quality scores, while our fifth proposed system (CFL-System 2) which is discussed in this section

employs the information of the head-pose for this purpose. This system which is published in [23] uses the superset of facial features explained in chapter 4, except the nose's feature. The methods for extraction these facial features are discussed in Section 4.2.

For scoring, this system uses a MLP similar to BFI-System 2, i.e. having extracted and normalized the ten quality measures for all the faces of the input video sequence, they are fed to the MLP to produce a quality score for each face. Neural networks can tackle problems that people are good at solving very well, like choosing the image with the higher quality. The networks are good at learning the features of the complex space of human faces. Furthermore, the behavior of neural networks in the cases that the input data is not complete is more reasonable than the simple combination of the normalized features as in [24]. It helps the system to work with the low-resolution videos where some of the facial features may not be extractable. The employed MLP has three layers with 10 neurons in its input layer, each corresponding to one of the extracted features, 4 neurons in the hidden layer and one neuron in the output layer indicating the quality score of the input face image. The method and data for training this neural network is the same as the system described in 5.3.1.1.

5.3.2.2.1 CFL-System 2: Experimental Results

Similar to the previous described systems, in this section, we first give an overview of the used databases and then the experimental results are given. For testing this system we have used four different databases. The first three databases are FRI CVL (DB3), Hermes (DB4) and AT&T (DB5) which have been explained in 5.3.1.1.1. The last database is:

Local Database (DB10): The sequences in this database are more realistic compared to the other databases. The 30 persons participating in this database are being asked to talk, change their gaze and head rotations while moving freely in front of a Logitech camera. More than 100 video sequences, each containing at least 150 images, have been captured from these people.

Having the quality score of each face image in a given video sequence, we use a three-step process to summarize the video sequence. Each step of this process tries to complete and evolve the result of the previous step. However, the result of each step can be used for a specific purpose in facial analysis systems.

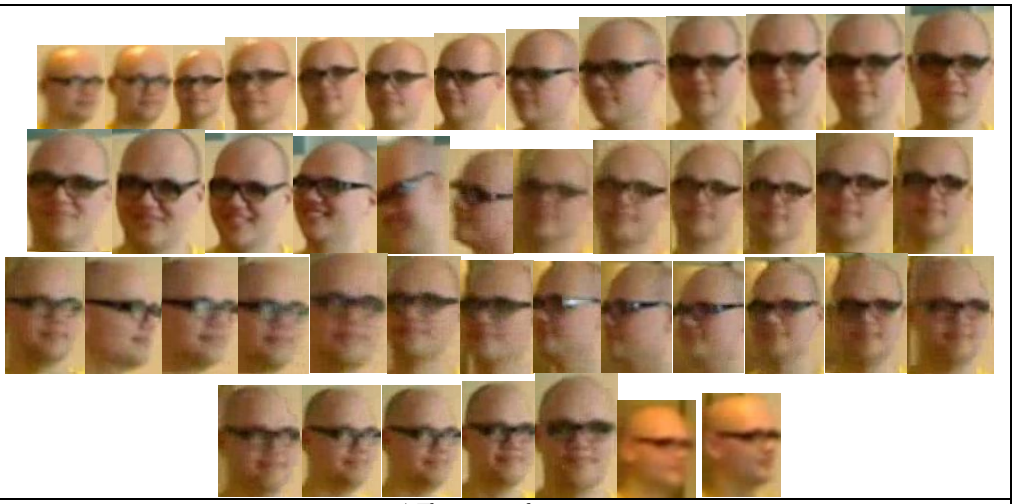
In the first step, we find the best face image of the input video sequence. The best face image of the sequence is defined as the image having the highest quality score from the MLP explained in the previous section. It is usually a frontal face image with a pan angle between -25 and 25 degrees (if there is such an image in the sequence). This image can for example be used for video indexing in huge databases of different video sequences. Figure 5-16(a) shows every mth frame ($3<m<15$) of a video sequence and Figure 5-16(b) shows its best face image(s), sorted by their quality score from left to right. Therefore, if the number of the faces in the input video sequence is n_1, this step reduces it to n_2. We have chosen $n_2=1$ for now,

however, it can be bigger for some specific applications as has been discussed in [8, 9, 12]. In cases like in [8, 9, 12], the output of this step contains the n_2-best face images of the video sequence, without taking care about the head-pose, as in Figure 5-16(b).

The second step tries to produce a complete face-log for the input video sequence. It means that the log should contain the best side-view face images (if any) as well as the best frontal face image. This face-log can be used as a concise representation of the input video sequence. To generate this complete face-log we first use the pan information to divide the input video sequence into three initial face-logs. Thus, if the number of the faces in the input video sequence is n_1, this step reduces it to n_2. Each of these initial face-logs corresponds to one direction: frontal face images, left and right side-view face images. The number of images (n_2) in each of these initial face-logs can be different. Figure 5-16(c1)-Figure 5-16(c3) show the three initial face-logs generated for the video sequence given in Figure 5-16(a). Then, using the explained quality measures each of these face-logs is reduced to n_3-best face images ($n_3<n_2$). These n_3-best images are the n_3 images with the highest quality score, and are denoted as intermediate face-log. For this step, we keep only the best face image for each log in each intermediate log, i.e. $n_3=1$.

The best frontal face image is the one with the best quality among its peers and least rotation, see Figure 5-16(c5). However, the best side-view images are the one with best quality and most rotation in pan, see Figure 5-16(c4) and Figure 5-16(c6). Therefore, to keep the generality of normalization formula in Equation 4-2, after obtaining the pan angle of each image, if it is outside of the range -25 to +25 degrees, the absolute value of the angle is reduced from 90 and the result is considered as the pan angle for the image. The complete face-log that is the output of this step composes these three intermediate face-logs. Having such a face-log from a video sequence is more than enough to identify the person, or even to make a 3D face model of that person (It will be discussed in the chapter 7).

The last step for evolving our face-log is to preparing it for a super-resolution algorithm. Such face-logs are denoted over-complete face-logs in this thesis. Super-resolution algorithms are a class of algorithms that are used for extracting one or more high-quality images from one or more low-resolution algorithms. These algorithms and producing an over-complete face-logs for them are discussed in chapter 6 and 7, respectively. Figure 5-16(d1)-Figure 5-16(d3) show over-complete face-logs obtained for the face-logs shown in Figure 5-16(c1) to Figure 5-16(c3), respectively. We will come back to this figure in chapter 7.

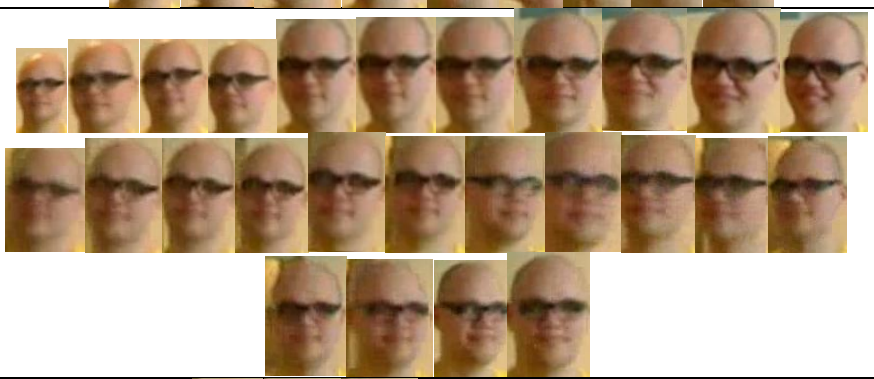


a) The input video sequence.

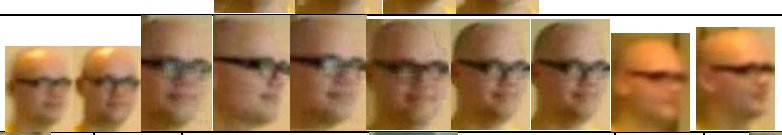
b) The best face images of the sequence without considering the head rotation.

c1

c2



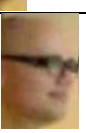
c3



c4

c5

c6



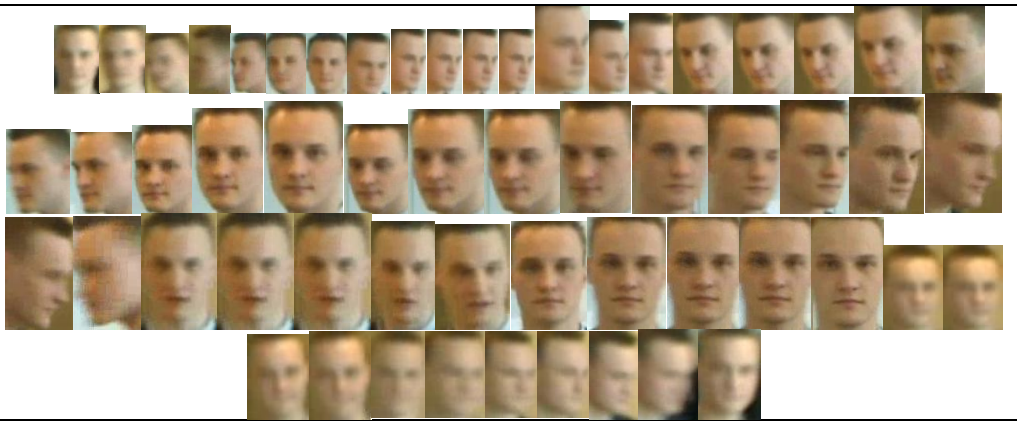
Continues to the Next Page

c) Constructing the complete face-log by dividing the input video sequence into three initial face-logs using the head pose information: c1) initial face-log of the right side-view images, c2) initial face-log of the frontal face images, c3) initial face-log of the left side-view images, c4) the best right side-view face image, c5) the best frontal face image, and c6) the best left side-view face image. The complete face-log is the composition of the last three best images.

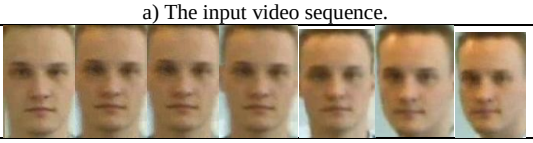
d) Constructing three over-complete face-logs for the super-resolution algorithm: d1, d2, and d3 are the best face-logs generated from c1, c2, and c3, respectively.

Figure 5-16: Summarizing an input video sequence to different face-logs.

Figure 5-17 shows another video sequence from DB10, and illustrates the process for summarization the input video sequence (given in part a of the figure) into face-log containing the best images of the sequence (shown in part b of the figure), the complete face-log (shown in part c), and finally the over-complete face-logs (shown in the last part of the figure). This figure will be discussed more in chapter 7.



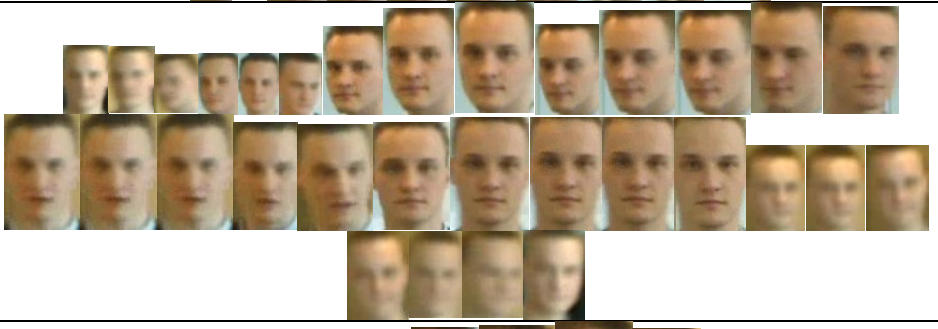
a) The input video sequence.



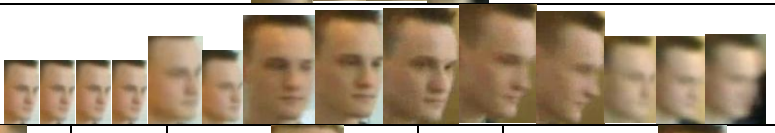
b) The best face images of the sequence

c1

c2



c3



c4

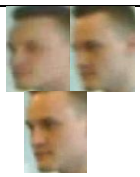
c5

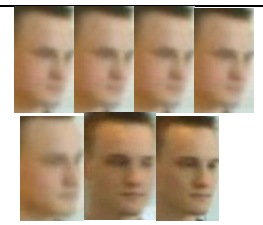
c6

Continues to the Next Page

c) Constructing the complete face-log by dividing the input video sequence into three initial face-logs using the head pose information: c1) initial face-log of the right side-view images, c2) initial face-log of the frontal face images, c3) initial face-log of the left side-view images, c4) the best right side-view face image, c5) the best frontal face image, and c6) the best left side-view face image. The complete face-log is the composition of the last three best images.

d1		d2		d3	





d) Constructing three over-complete face-logs for the super-resolution algorithm: d1, d2, and d3 are the best face-logs generated from c1, c2, and c3, respectively.

Figure 5-17: Summarization of another input video sequence to different face-logs.

5.4 Summary

This chapter of the thesis introduces the concept of face-log. The close relationship between the *scoring* step of the proposed face quality assessment systems and face-log construction is illustrated in this chapter. Three different types of face-logs have been introduced. The first one contains the best face image(s) of the input video sequence and it can be used for example in video indexing. The second type of face-logs discussed in this chapter is complete face-log, which can be used as a complete and concise representation of a video sequence. Different methods for constructing these two types of face-logs are discussed in this chapter. However, discussing the construction of the last type of face-logs, over-complete face-logs, is postponed to chapter 7. The purpose of constructing over-complete face-logs is using them in super-resolution algorithms. These algorithms are discussed in the next chapter.

References

[1] P. Griffin, "*Understanding the Face Image Format Standards*," National Institute of Standards and Technology Workshop, American National Standards Institute, Gaithersburg, Maryland, USA, 2005

[2] A.N. Zamani, M.K. Awang, N. Omar, and S.A. Nazeer, "*Image Quality Assessments and Restoration for Face Detection and Recognition System Images*," IEEE 2nd Asia International Conference on Modeling & Simulation, Kuala Lumpur, Malaysia, 2008.

[3] Q. Xiong and C. Jaynes, "*Mugshot Database Acquisition in Video Surveillance Networks Using Incremental Auto-Clustering Quality Measures*," IEEE Conference on Advanced Video and Signal-based Surveillance, Miami, USA, 2003.

[4] H. Fronthaler, K. Kollreider and J. Bigun, "*Automatic Image Quality Assessment with Application in Biometrics*," IEEE Conference on Computer Vision and Pattern Recognition Workshop, New York, USA, 2006.

[5] Y. Rodriguez, F. Cardinaux, S. Bengio and J. Mariethoz, "*Estimating the quality of face localization for face verification*," IEEE International Conference on Image Processing, Singapore, 2004.

[6] G. Xiufeng, Z. Stan, L. Rong, and P. Zhang, "*Standardization of Face Image Sample Quality*," International Conference on Advances in Biometrics, Seoul, Korea, 2007.
[7] M. Subasic, S. Loncaric, T. Petkovic and H. Bogunvoic, "*Face Image Validation System*," 4th International Symposium on Image and Signal Processing and Analysis, Zagreb, Croatia, 2005.
[8] A. Fourney and R. Laganiere, "*Constructing Face Image Logs that are Both Complete and Concise*," 4th IEEE Canadian International Conference on Computer Vision and Robot Vision, Canada, 2007.
[9] K. Nasrollahi and T.B. Moeslund, "*Complete Face-logs for Video Sequences using Quality Face Measures*," IET Journal of Signal Processing, vol. 3, no. 4, pp. 289-300, 2009.
[10] G. Zhang and Y. Wang, "*Asymmetry-Based Quality Assessment of Face Images*," Lecture Notes in Computer Science, Advances in Visual Computing, vol. 5876, pp. 499-508, 2009.
[11] A.D. Bimbo, F. Dini and G. Lisanti, "*A Real Time Solution for Face-logging*," the 3rd IEEE International Conference on Crime Detection and Prevention, pp. 1-6, 2009.
[12] K. Nasrollahi and T.B. Moeslund, "*Face Quality Assessment System in Video Sequences*," B. Schouten, N.C. Juul, A. Drygajlo and M. Tistarelli.(eds.) Biometrics and Identity Management, Springer Lecture Notes in Computer Science , vol. 5372, pp. 10-18, Springer Verlag Berlin Heidelberg, 2008.
[13] F. Solina, P. Peer, B. Batagelj, S. Juvan, J. Kovac, "*Color-Based Face Detection in the "15 Seconds of Fame Art Installation*," Conference on Computer Vision / Computer Graphics Collaboration for Model-based Imaging, Rendering, Image Analysis and Graphical special Effects, pp. 38-47, France, 2003.
[14] Hermes project (FP6 IST-027110): http://www.cvmt.dk/projects/Hermes/index.html
[15] K. Nasrollahi and T.B. Moeslund, "*Real Time Face Quality Assessment for Face-log Generation*," International Conference on Machine Vision, Image Processing, and Pattern Analysis, Bangkok, Thailand, 2009.
[16] AT&T Laboratories Cambridge, Face Database, available at: http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html
[17] Collection of Facial Images: Faces96, available at: http://cswww.essex.ac.uk/mv/allfaces/faces96.html
[18] E.H., Mamdani and S. Assilian, "*An Experiment in Linguistic Synthesis with a Fuzzy Logic Controller*," International Journal of Man-Machine Studies, vol. 7, no. 1, pp. 1-13, 1975.
[19] J.C. Bezdek, J.M. Keller, R. Krishnapuram, and N.R. Pal, "*Fuzzy Models and Algorithms for Pattern Recognition and Image Processing*," Kluwer Academic Publishers Norwell, Massachusetts, USA, 2005.
[20] D. Roth, P. Doubek, and L.V. Gool, "*Bayesian Pixel Classification for Human Tracking*," IEEE Workshop on Motion and Video Computing, Breckenridge, Colorado, USA, 2005.
[21] P. Viola, and M.J. Jones, "*Robust Real Time Face Detection*," Internatioal Journal of Computer Vision, vol. 57, no. 2, pp. 137-154, 2004.
[22] H. Bibby and I. Reid, "*Robust Real-Time Visual Tracking using Pixel-Wise Posteriors*," 10th European Conference on Computer Vision, Marseille, France, 2008.
[23] Kamal Nasrollahi, Thomas B. Moeslund, and Mohammad Rahmati, "Summarization of Surveillance Video Sequences Using Face Quality Assessment," To appear in International Journal of Image and Graphics, January 2011.
[24] M. Sordo, "Introduction to Neural Networks in Healthcare," OpenClinical, Knowledge Management for Medical Care, Harvard, 2002.

CHAPTER 6

SUPER-RESOLUTION: A LITERATURE SURVEY

6 Chapter 6: Super-Resolution: A Literature Survey

6.1 Introduction

This chapter is devoted to a survey on super-resolution techniques in the literature. Super-resolution is a process for obtaining one or more high-resolution image(s) from one or more low-resolution input image(s) [1-389]. It seems that the works published in [1-3] are the pioneer works in the field of super-resolution. Gerchberg [1] in 1974 used the term super-resolution to introduce his iterative method for continuing a given spectrum beyond its given limits and therefore achieving a resolution beyond the diffraction limits. Then, Santis and Gori [2] used prelate spheroidal wave functions to extend Gerchberg's method. They showed that a perfect imagery can be achieved in the limit of an infinite number of iterations. Fiernup [3] then proposed a method for reconstruction of a general object from the modulus of its Fourier transform. He mentioned that this technique can be useful for obtaining high-resolution imagery from interferometer data.

Super-resolution has a wide range of applications in different fields of science from improving the quality and resolution of astronomical images acquired by huge telescopes to enhancing the quality of images captured by tiny surveillance cameras. This process has been widely employed in facial analysis systems, medical image processing, data transmission, plate reading, digital holography, aviation safety in extremely poor visibility conditions, and many other constantly increasing applications.

A high-resolution image has a higher number of pixels and therefore more visible details of the captured scene than a low-resolution image. The value of the software approach for obtaining a high-resolution image (super-resolution algorithms) becomes clearer when we elaborate the hardware solutions of the problem: increasing the chip size and/or decreasing the pixel size in order to have more pixels of a given scene. Increasing the chip size in order to cope with the need of very high-resolution images is an expensive solution. Furthermore, it leads to an increase in the capacitance. Large capacitance causes problems in the speed of charge transfer rate [142]. Thus, this method is not used that much. The second hardware approach, reducing the pixels size, decreases the amount of light and energy that each element of a CCD or CMOS, corresponding to each pixel, can collect [142]. The reduction in the collected light obviously results in the shot noise and subsequently quality degradation. Thus, there is a lower bound limitation for the pixel size and it cannot be smaller than this limit. The current technology has almost reached this limit and further improvement using this method can hardly be achieved [142].

Limitations of the hardware approaches for increasing the spatial resolution of imaging devices, and innumerous applications and preference of high-resolution images have made the super-resolution algorithms a very interesting field for the researchers. Consequently, there are so many publications on super-resolution algorithms in the literature. Thus, there is a need for

grouping different ideas related to super-resolution to help new researchers finding their ways in this everyday growing field. This is exactly the purpose of this survey.

There are some other publications in the literature [63, 64, 141, 142, 156, 191, 208, 223, and 237] surveying different types of super-resolution techniques. Among them [63] and [142] provide a better understanding of different types of super-resolution algorithms, their different steps, and their cons and pros. However, they are almost old and don‘t cover an emerging group of super-resolution algorithms as is discussed in this thesis. Furthermore, they do not include the latest achievements in the classical super-resolution algorithms. This chapter of the thesis performs a comprehensive and meanwhile, as much as possible, concise survey starting from 1974, going through 90s papers and being focused on the papers published in the last decade. That section of this paper which is devoted to reviewing the old papers is mostly based on the references [63] and [142]. These references have been used thoroughly for preparing that section.

The rest of this chapter is organized as follows: the next section introduces a new grouping schema for the super-resolution algorithms. Based on this grouping the super-resolution algorithms are divided into three classes. Therefore, the next three sections describe each of these classes. Finally, the chapter is drawn to conclusion in the last section.

It should be mentioned that the sections describing the three classes are not balanced. This is due to the number of papers found for each group and also the novelty of the last two parts.

6.2 Grouping Super-Resolution Algorithms

Super-resolution algorithms can be classified in many different ways using different factors. For example based on the number of the inputs (single image or multiple images), the observation model, the method they use for restoration, the regularization method, the domain in which the process is carried out (the frequency or spatial domains), etc. The schema that is used in this chapter for grouping different super-resolution algorithms is shown in Figure 6-1. This schema divides the super-resolution algorithms into three different classes: reconstruction-based, recognition-based and hybrid methods. The previous surveys on super-resolution algorithms mainly cover the reconstruction-based. However, they do not consider the recognition-based ones. The reason for this ignorance maybe is that at the time of preparing these surveys there were not enough papers on the recognition-based methods. Nowadays, especially in the last few years, these types of super-resolution algorithms have gotten a great interest among the researchers. In addition to their own potentials, their combination with reconstruction-based methods has emerged a new branch of super-resolution algorithms that are denoted in this thesis as hybrid methods. These methods are not covered in any previously published surveys on super-resolution and are the main difference between this chapter of the thesis and those distinguished surveys [63, 142] . The following sections describe each of these groups of algorithms.

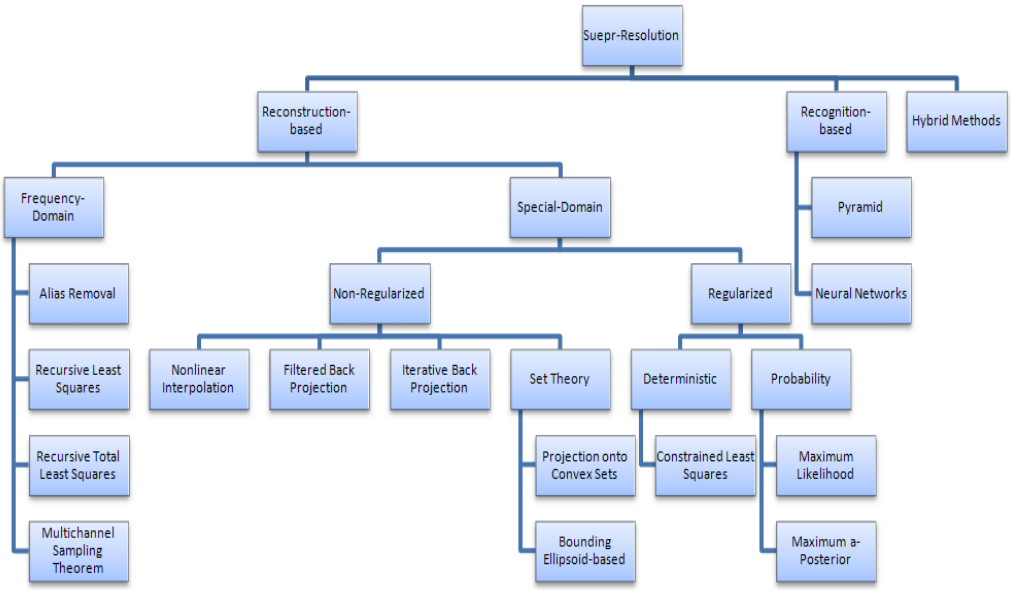


Figure 6-1: The introduced schema for grouping super-resolution algorithms.

6.3 Reconstruction-based Super-Resolution Algorithms

Even though [63] reviews the reconstruction-based super-resolution algorithms very well, it does not cover the latest achievements in this filed. Some of these achievements are then included in [141] and [142]. However, in order to keep the generality of this chapter, reorganizing this group of super-resolution algorithms and including the latest state-of-the-art in the field of reconstruction-based super-resolution algorithms, we have covered this group. It should be mentioned that the grouping schema of the classical reconstruction-based super-resolution algorithms are adopted from [63, 141, and 142].

Reconstruction-based super-resolution algorithms typically need more than one low-resolution input image. These algorithms usually employ an observation model. It is considered that the low-resolution input images are obtained from the original high-resolution scene using this observation model. Different observation models are proposed for reconstruction-based super-resolution algorithms in the literature [7, 12, 20, 25, 28, 33, 36, 39, 62, 82, 83, 95, 96, 106, 108, 113, 162, and 380]. The most important observation model that is widely used among the researchers and its different parts are explained in the next section.

6.3.1 Observation Model (Imaging Model)

Based on the most common observation model (Figure 6-2) it is considered that the available low-resolution input images are obtained from the high-resolution original scene by warping,

blurring and down sampling the scene. Considering H as the high-resolution original image of the scene and X_i, $i=\{1...n\}$ as the n low-resolution input images, the imaging model is defined as:

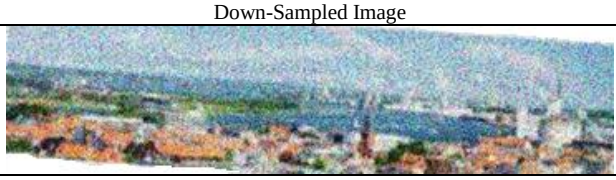
$$X_i = DB_iW_iH + e_i \qquad 6\text{-}1$$

Orignal High-Resolution Scene

Wrapped Image

Blurred Image

Down-Sampled Image



Introducing Noise into the Imaging Model

Figure 6-2: The imagong model under which the low-resolution images (bottom of the image) are considred to be obtained from the high-resolution scene (top of the image).
The orignal image (top raw) is taken from www.aku-aalborg.dk

where, D is the down sampling matrix, B_i and W_i are the blurring matrix and warping matrix for producing the ith low resolution image from the high-resolution image, respectively. The imaging error that can be introduced into the imaging model during the imaging process is modeled by e_i. Figure 6-2 shows this imaging model.

Early reconstruction-based super-resolution algorithms did not have the ability to cope completely with such a model and therefore had difficulties in real-world applications. The state-of-the-art super-resolution algorithms, however, are fully consistent with such models and hence widely employed in the real-world scenarios. Following the model showing in Figure 6-2, in order to reconstruct the super-resolved high-resolution image a geometrical and a photometrical registration should be applied to the wrapped and blurred low-resolution input images, respectively. These steps can be implemented separately or simultaneously according to the adopted reconstruction methods [142]. Before going into the details of different reconstruction-based super-resolution algorithms the different steps of the imaging model are reviewed in the following subsections.

6.3.1.1 Geometric Registration

The low-resolution input images to reconstruction-based super-resolution algorithms need to have sub-pixel relative misalignment. These sub-pixel misalignments are the facts behind the scene which enable the reconstruction-based algorithms to super-resolve the missing high-resolution details (See Figure 6-3). If the relative misalignments between the low-resolution input images are not fractions of pixels, the restoration algorithm cannot produce any new information for the high-resolution output image.

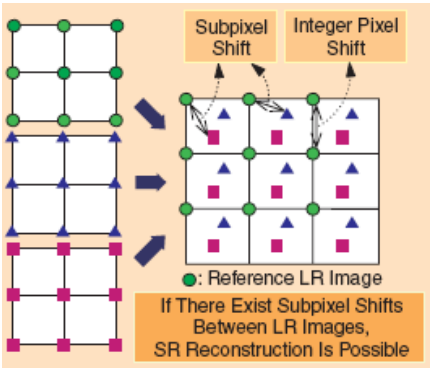


Figure 6-3: Relative sub-pixel misalignments between low-resolution input images (left) and their effect on the super-resolved high-resolution image (right). The image is taken from [142]

In order to obtain the information about the low-resolution sub-pixel misalignments, vast majority of multi-frame super-resolution algorithms employ a registration process.

Registration, in other words, is finding a transformation which relates the low-resolution input images with each other. The most general case of transformation is the 8 degree of freedom. This is called planar projective transformation or homography. For special camera motions the mapping between images of planes is simpler than a general homography. These simpler mappings have fewer degrees of freedom, and consequently can often be more reliably and rapidly estimated [98].

If the camera motion consists of a translation parallel to the image plane and rotation about the principal axis then images are related by 4 degrees of freedom, *similarity transformation*. This motion is often the case in satellite imaging and document scanning. Under telephoto viewing conditions, where perspective effects such as the convergence of imaged parallel lines are negligible, images are related by 6 degrees of freedom *affine transformation* [98].

Both similarity and affinity mappings are perfectly correct and valid transformations under certain imaging conditions they are subgroups of a planar homography. By contrast, some authors have worked with more general imaging conditions in which only the full homography can correctly capture the image transformation, but for reasons of numerical stability and/or simplicity of optimization have chosen to approximate the homography by a Taylor expansion to second-order, resulting in the *biquadratic transformation*. This has 12 degrees of freedom, but is unable to correctly model perspective effects [98].

There are many methods for image registration for the purpose of super-resolution [7, 19, 157, 171, 226, 227, 235, 236, 264, and 281]. A typical registration algorithm finds the features of interest in the low-resolution images and then uses some methods to estimate point correspondences and compute homographies between the images [262]. Irani and Peleg [14] estimate the registration parameters using an iterative technique which wraps the low-resolution images into each other's frames. Their method considers a vertical and a horizontal shifts and a rotation between the low-resolution images (This method is further discussed in the next chapter). Hardie et al. [54, 55] employ a joint MAP registration and high-resolution image estimation using a sequence of under sampled images. They used the estimation of the super-resolved image to refine the registration of the low-resolution images. Villena et al [171] uses a simultaneous auto-regression (SAR) algorithm to find the unknown displacements between the low-resolution inputs. Borman et al. [157] considers a linear observation model to multi-frame super-resolution restoration under conditions of non-affine image registration. Patanavijit and Jitapunkul [226] use an affine block based registration. Pickup et al. [262-264] perform Bayesian integration over the unknown image registration parameters, deriving a cost function whose only variables of interest are the pixel values of the super-resolution image. Even though most of the authors have tried to reduce the registration errors, Costa and Bermudez [326] show that a certain amount of registration error may, in fact, be beneficial for the performance of the least mean square super-resolution reconstruction adaptive algorithm.

6.3.1.2 Photometric Registration

By analogy with the geometric registration which obtains the geometric degradation information that are imposed on the low-resolution input images during the imaging process, the photometric registration tries to find the degradation that are imposed on the low-resolution images during the imaging process as a result of the photometric factors [98, 99]. Methods for photometric registration can be broadly divided into two groups: in the first group which is more realistic the effects of the camera adjustment and the illumination changes on the low-resolution images are considered. In this group, the photometric information is obtained for each image, separately. In the second group which is more common, the photometric degradations are only considered to be a result of the camera's point spread function and hence it is the same for all the low-resolution input images (as in Figure 6-2). Therefore, the problem of photometric registration in the second group of algorithms is changed to the problem of finding the point spread function of the camera. This point spread function can be decomposed into factors representing the blurring caused by camera optics and the spatial integration performed by a CCD sensor [83].

Different optical systems coupled with sensors of various geometries lead to a variety of possible shapes for point spread functions. Though it is a hard to estimate this function, many methods have been studied to do so. Nguyen et al. [79, 106 and 107] generalized cross validation approach to learn the blur parameters. Tipping and Bishop [125] and then Abad et al. [128] used maximum likelihood for this purpose. Generally, the point spread functions approximated by a simple parametric function centered on each low-resolution pixel. The two most common ones are an isotropic 2D Gaussian or a circular disk (top-hat function) [262]. However, the methods for obtaining the point spread function can be generally divided into the following groups [114]:

- Use camera manufacture information which is usually hard to get.
- Analysis a picture of a known object [20, 33]
- Blind estimation of the point spread function from the images [82].

6.3.1.3 Noise in the Imaging Model

The noise in the imaging model of the reconstruction-based super-resolution algorithms is usually considered to be an additive *i.i.d* Gaussian noise.

6.3.2 Reconstruction Process

As it is shown in Figure 6-1 the reconstruction-based super-resolution algorithms are generally divided into two groups: frequency domain and spatial domain methods. Frequency domain reconstruction-based super-resolution algorithms are the ones which were developed first. Even though development of these algorithms started form 1974 [1], the most well-known frequency

domain reconstruction-based super-resolution was introduced by Tsai and Huang [7] ten years later. This algorithm and the other frequency domain reconstruction-based super-resolution algorithms are reviewed in the next subsection.

6.3.2.1 Frequency domain

Frequency domain approaches [1, 3, 7, 16, 24, 27, 29, 30, 52, 149, 152, 198, 235, 236, 291, and 382] are divided into four groups [63]. These groups are very briefly discussed in the following subsections. For more detailed survey of frequency domain approaches the reader is referred to [63].

6.3.2.1.1 Alias Removal

The work introduced by Tsai and Huang [7] belongs to this group and it is based on the formulation of the restoration problem in frequency domain. This approach uses the shifting property of the Fourier transform to model global translational motions between its low-resolution input images. Furthermore, the sampling theory is used to perform the restoration from the low-resolution input images. This is considered as super-resolution algorithm as it tries to recover the information beyond the Nyquist limit. Employing the shifting property of Fourier transform and also sampling theorem are the common characteristics of most of the frequency domain approaches. The shifting property of Fourier transform is only capable of coping with translational motion between low-resolution input images. Therefore, these approaches are not suitable to be used with imaging models like the one shown in Figure 6-2 in which any kind of motions can be present among the low-resolution input images. However, the method in [7] was developed to work with satellite images in which the only available motions are translational. Thus, this method was suitable for that purpose. Though, the blurring effects resulted from the satellite motion and the effects of noise on the imaging process were both ignored in that work. The latest two restrictions were later addressed in [24]. Tekalp et al. [52] later developed frequency domain systems that were able to cope with more motion models [63].

6.3.2.1.2 Recursive Least Squares

The methods of this group [16] are an extension of the basic work of Tsai and Huang [7] in which more general motion models than the translational motion and noise involvement are considered. Furthermore, a computational efficient recursive least square method has been employed for reconstruction the high-resolution images from the low-resolution aliased input images [63].

6.3.2.1.3 Recursive Total Least Squares

The methods of this group [27, 29] are extensions of the previous group in which the degree of robustness against the error has been improved using a recursive total least square algorithm [30, 63].

6.3.2.1.4 Multichannel Sampling Theorem

The outstanding work of this group is the one by Ur and Gross [25]. They use the multichannel sampling theorem to reconstruct the high-resolution output signal by passing the low-resolution input signals through a set of mutually independent filters, summing the output and performing a band limited interpolation [63].

Even though, the frequency domain approaches are really simple and computationally efficient, they have severe problems: they are mostly limited to global translational motion models and are not that flexible to be used with more general motion models. Furthermore, they cannot cope completely with degradation models such the one shown in Figure 6-2. The last and the most serious problem is that they cannot include the a-prior knowledge of the problem into their formulations [63]. Therefore, the spatial domain approaches are developed in which all the problems of the frequency domain methods are dealt with.

6.3.2.2 Spatial domain

Despite the fact that the frequency domain approaches were the ones which were developed first for dealing with the super-resolution problem, they are less used than the spatial domain methods, nowadays. This is due to their aforementioned problems. These problems are very well dealt by developing the spatial domain methods for the problem. The spatial domain algorithms [4, 6, 5, 8, 10, 11, 13, 14, 12, 16, 20, 21, 18, 28, 41, 47, 48, 51, 52, 55-57, 59, 60, 67, 67, 72, 80, 85, 72, 80, 202, 218, 224, 231, 247, 252, 265, 281, 332, 348, and 363] more or less have the ability to cope with the observation model of the imaging systems. Furthermore, they have the ability to include the a-priori constraints of the problem domain into their formulation. Inclusion the a-priori constraint is of critical importance for super-resolution algorithms as most of them are in the form of an inverse problem. Inverse problems are generally ill-posed and ill-conditioned. In spatial domain super-resolution algorithms usually the a-priori constraints are utilized as generalization terms to convert the ill-posed problem to a well-posed. Depending on using the regularization term the spatial domain approaches for super-resolution are generally divided into two groups. The descriptions of these groups are as follows.

6.3.2.2.1 Non-Regularized

6.3.2.2.1.1 Nonlinear Interpolation

In the algorithms of this group [4, 5, 8, 10, 11, 12, 16, and 18] usually the low-resolution input-images are registered to a reference image. The result is a composite image of non-uniformly spaced samples. Then, a high-resolution grid is considered for the high-resolution output image and the points of the non-uniformly spaced samples of the composite image are interpolated and resampled to that high-resolution grid [63].

6.3.2.2.1.2 Filtered Back Projection

The methods of this group [10] are inspired by computer aided tomography. The reconstruction problem in this group is formulated as a liner inverse problem. Then, a back projection operator is used to solve the problem. Rewriting Equation 6-1, to a simpler form we have [63]:

$$X = MH \qquad 6\text{-}2$$

in which X is the stack of low-resolution input images, M is the observation model matrix, and H is the high-resolution output image. The reconstruction methods of this group of algorithms are usually in the form of a back-projection of the observed low-resolution images as [63]:

$$\bar{H} = M'X \qquad 6\text{-}3$$

from 6-2 $\bar{H} = M'MH$

in which M' is the back projection matrix. Usually a fast Fourier domain inverse filtering for $M'M$ can be applied to obtain the back projected observed data:

$$\bar{H} = (M'M)^{-1}M'X \qquad 6\text{-}4$$

These can be then solved using techniques like Linear Minimum Square Errors [63].

6.3.2.2.1.3 Iterative Back Projection

The ideas of the algorithms [6, 11, 14, 20, and 28] of this group are generally similar to the previous group. However, the main difference is that usually an iterative approach is involved. In each iteration, the difference between the low-resolution images (obtained from the estimated high-resolution image) and the original low-resolution images is considered as a residual error and will be reflected in the next iteration for estimating the high-resolution image. This process is repeated until some convergence is achieved [63].

Conforming the super-resolution result to the low-resolution input images, employed by iterative back projection methods, is an interesting idea. However, there are some concerns regarding the methods used for measuring the error, existence of the solution, its uniqueness

and its stability. Furthermore, it is almost impossible to involve the a-prior information in these algorithms [63].

6.3.2.2.1.4 Set Theory

Similar to the iterative back projection methods, set theory methods [13, 21, 41, 47, 48, 51, 52, 59, 60, and 67] use an iterative method for obtaining the high-resolution image. However, they have the ability to incorporate a-prior knowledge about the solution into the reconstruction step [142]. The inclusion of the a-prior knowledge forces the reconstructed super-resolved solution to be consistent with the low-resolution input images. Furthermore, a set of constraints are usually defined in these methods which candidate solutions should be consistent with them. The most common constraint is the smoothness of the response. The employed iterative methods should simultaneously satisfy all the constraints [63].

6.3.2.2.1.4.1 Projection onto Convex Sets

In the algorithms [13, 21, 52, 59, 60, and 67] of this group the constraint sets that the super-resolved image should be consistent with are convex sets. It means that the constraints are defined as convex sets in a vector space which contains all the possible super-resolved images. In another words, the inclusion of the a-prior knowledge can be considered as restricting the super-resolved image to be a member of a closed convex set C_i which are defined as a set of vector satisfying a specific property [13]. The super-resolved image is a convex set which belongs to the intersection of all the constraints: $C_H = \cap_{i=1}^{nc} C_i$. In which C_H is a convex set which is equal to the intersection of all the convex set, any vector in this space can be considered as a super-resolved image (H), and nc is the number of convex sets. In order to find the high-resolution output image a recursive process is considered as follows [63]:

$$H^{z+1} = P_{nc} P_{nc-1} \dots P_1 H^z \qquad 6\text{-}5$$

which H^0 is an arbitrary starting point, and P_i is a projection operator which projects an initial guess of the high-resolved image onto the closed, convex sets, C_i, $\{i=1,\dots, nc\}$. The first algorithms of this group [13] were not able to completely cope with imaging model [63]. However, this problem is dealt with in the later proposed systems [52, 59, 60, and 67].

6.3.2.2.1.4.2 Bounding Ellipsoid-based

The algorithms [41, 47, 48, and 51] of this group are a different version of projection onto convex sets in which an ellipsoid is employed to constraint the sets. Given a set of ellipsoidal constraint sets, a bounding ellipsoid is computed. The centroid of this ellipsoid is taken as the super/resolution image estimate [63].

The set theory methods are generally simple, have the ability to cope with the imaging model and can incorporate the a-prior information into their formulation. However, they suffer from severe problems: first of all the super-resolved image obtained by this method is not necessarily unique. The final output of the system highly depends on the initial guess. Besides, the computational costs of these algorithms are high and their convergence rate is slow. Furthermore, implementing the required projection operators is not always easy [63, 142].

Most of the problems of the set theory methods are dealt with in the super-resolution reconstruction-based methods that use a regularization term. Following is the description of this group of algorithms [63].

6.3.2.2.2 Regularized

In contrast to the previously mentioned methods, most of the reconstruction-based super-resolution algorithms deal with this problem as an inverse problem [21, 54-57, 65, 72, 80, 85, 202, 218, 224, 231, 247, 252, 265, 281, 332, 348, and 363]. It means that having a set of low-resolution input images they try to compute H (See Equation 6-1) by inversing the imaging process (See Figure 6-2). If enough numbers of low-resolution input images (X_i) are available for reconstructing the high-resolution image, obtaining H is a well-posed problem. If enough numbers of low-resolution images are not available, then computing H from Equation 6-1 is an ill-posed problem. In the later (which is usually the case in the real-world applications), we need a regularization term to convert the ill-posed problem into a well-posed one. Different approaches for solving this problem and different regularization terms are discussed in the following subsections [141].

6.3.2.2.2.1 Deterministic

This group of reconstruction-based super-resolution algorithms [55, 56, 57, 65, and 85] can completely cope with the imaging model shown in Figure 6-2. One of these algorithms is explained in the following subsection.

6.3.2.2.2.1.1 Constrained Least Squares

Following Equation 6-1, constraint least squares can be formulated as [21]:

$$\sum_{i=1}^{n} ||X_i - DB_iW_iH||^2 + \alpha||CH||^2 \qquad 6\text{-}6$$

which $\|.\|$ is the l_2-norm, C is a high-pass filter which constraints the high-resolution image H to be smooth, $||CH||^2$ is the regularization term and α is the regularization parameter. This parameter $\propto$ controls the tradeoff between the fidelity of the super-resolved response to the

low-resolution input images and its smoothness [142]. A unique estimation of *H*, the super-resolved response, in the cost function of 6-6 can be found by for example an iterative technique:

$$\left[\sum_{i=1}^{n}[DB_iW_i]^T[DB_iW_i] + \alpha C^T C\right]\bar{H} = \sum_{i=1}^{n}[DB_iW_i]^T X_i \quad 6\text{-}7$$

From this, we have:

$$\bar{H}^{z+1} = \bar{H}^z + \beta\left[\sum_{i=1}^{n}[DB_iW_i]^T[X_i - DB_iW_i\bar{H}^z] - \alpha C^T C\bar{H}^z\right] \quad 6\text{-}8$$

which β is the convergence parameter and *T* is the transpose operator [21, 142].

6.3.2.2.2.2 Probability

Probability based approaches [54, 72, 80, 202, 218, 224, 231, 247, 252, 265, 281, 332, 348, and 363] are generally divided into two groups: Maximum Likelihood and Maximum a-Posterior. These approaches have the power of complete consideration of the a-prior knowledge and the details of the imaging model. Probability based reconstruction-based super-resolution algorithms usually follow a Bayesian formulation. Therefore, having the imaging model shown in Figure 6-2, they can be formulated as:

which γ is the inverse of the standard deviation of the proposed Gaussian model and *N* is the

$$p(X_i|\{H, B_i, W_i\}) = \left(\frac{\gamma}{2\pi}\right)^{\frac{N}{2}} e^{-\frac{\gamma}{2}(\|X_i - B_iW_iH\|_2^2)} \quad 6\text{-}9$$

number of the pixels of the low-resolution input images.

6.3.2.2.2.2.1 Maximum Likelihood

The Maximum Likelihood solution of the super-resolution problem is the super-resolved image which maximizes the probability of having the low-resolution input images:

$$\bar{H}_{ML} = argmax_H(p(X_i|\{H, B_i, W_i\})) \quad 6\text{-}10$$

Maximum Likelihood solutions are found to be ill-conditioned in real-world applications. It means that they are not stable with regards to the high-frequency oscillation in the low-resolution input images. This problem is dealt with in Maximum a-Posterior methods.

6.3.2.2.2.2.2 Maximum a-Posterior

Maximum a-posterior approaches for reconstruction-based super-resolution algorithms are obtained following a Bayes theorem formulation of the problem:

$$p(H|\{X_i, B_i, W_i\}) = \frac{p(X_i|\{H, B_i, W_i\})p(H)}{p(X_i|\{B_i, W_i\})} \quad 6\text{-}11$$

The maximum a-posterior can then be found by ignoring the constant denominator and maximizing the numerator with respect to the high-resolution image *H*:

$$\bar{H}_{MAP} = argmax_H(p(X_i|\{H, B_i, W_i\})p(H)) \quad 6\text{-}12$$

The details of such algorithms can be found in [63, 142] and also in the next chapter of this thesis. Many different regularization terms have been used in the literature. Most of these terms have Gaussian forms and impose a smoothness constraint on the high-resolution output. Hardie et al. [54, 55] use the l_2-norm of a Laplacian filter for this purpose. Huber Markov Random Fields [21, 56, and 57], Gaussian Markov Random Fields [111, 184, and 366], Bilinear Total Variation [134, 162-165], Gibbs, and Lorentzian-Tikhonov [260] are among the other registration terms that are used in the literature.

6.4 Recognition-based Super-Resolution (Hallucination) Algorithms

The second large group of super-resolution algorithms is the recognition-based ones, also called Hallucination algorithms or example based super-resolution algorithms [9, 70, 97, 118, 144, 150, 161, 168, 186, 192, 195, 196, 199, 216, 232, 247, 254, 255, 269, 270, 272, 277, 290, 306, 311, 333, 335, 336, 340-343, 354, 358, 360, 361, 367, 377, 378, 379, and 385]. In contrast to the reconstruction-based super-resolution algorithms which require multiple low-resolution images of the same scene as input, the recognition-based super-resolution algorithms can work with only one image. This is shown in Figure 6-4.

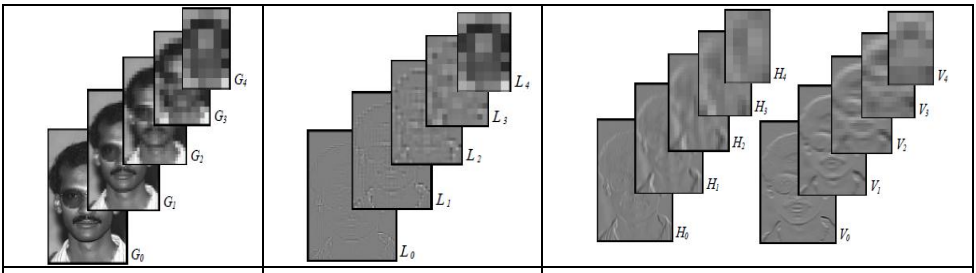
Figure 6-4: Recognition-based (left) vs. Reconstruction-based (right) super-resolution algorithms

The recognition-based super-resolution algorithms are usually application dependent, for example they are developed for face, text, plate-reading, etc. The first recognition-based super-resolution algorithm was first developed by Mjolsness [9] for hallucination of finger print images. The recognition-based super-resolution algorithms usually have a training database. For preparing the training database a set of high-resolution images of the object of interest are obtained then they are down-sampled to produce their low-resolution corresponding images. In the training step a learning algorithm is employed to learn the relationship between the low-resolution images and their corresponding high-resolution ones. Having a low-resolution input image, the system finds the closest image in the training database and uses the relationship between this low-resolution image and its corresponding high-resolution image to hallucinate the missing high-resolution details of the input image.

Even though, this group of algorithms introduced in 1985, they got interest in the last decade. The most successful hallucination algorithms are the ones developed for face images. Hallucination algorithms can be either characterized by the algorithm or model they use for learning, like hidden Marko models, neural networks, etc. or by the method they construct the training database. Constructing training database using pyramid-based methods and using neural network algorithms for learning step are explained in the following subsections.

6.4.1 Pyramid-based

In the pyramid-based recognition-based super-resolution algorithms the training databases contain a set of pyramids of the training images. These pyramids can be Gaussian, Laplacian, or derivate (features) pyramids in different resolutions (See Figure 6-5) or can be built by even more sophisticated features [19, 70].



Gaussian Pyramid	Laplacian Pyramid	Derivare Pyramids

Figure 6-5: Differen pyramids used in a recognition-based super-resolution system for face images [70].

6.4.2 Neural Networks-based

In the algorithms [9, 109, 124, 140, 146, 188, 257, 271, 304, and 323] of this group, different typed of neural networks are employed to hallucinate the missing high-resolution details of the low-resolution inputs. In most of such systems the neural networks take the pixels' values of the low-resolution image as its inputs while in others neural networks are used with for example wavelets of the low-resolution input image.

6.5 Hybrid Super-Resolution Algorithms

Most of the above mentioned super-resolution algorithms have their own pros and cons. Therefore, in order to provide new super-resolution systems, different authors have tried to combine the previously mentioned methods with each other. This group of hybrid algorithms can be divided into two classes: in the first class, different spatial domain reconstruction-based super-resolution algorithms are combined together into hybrid systems. These systems are reviewed very well in [63]. In the second class of hybrid algorithms, which have recently emerged, different reconstruction-based and recognition-based systems are combined together into hybrid systems.

The improvement factor of reconstruction-based super-resolution algorithms is practically limited to factors close to two [328]. Though, the improvement factor of the recognition-based super-resolution algorithms depends on the relationship between the high-resolution and their corresponding low-resolution images in their training database. High resolution images in these training databases of these algorithms are usually of double size of their corresponding low resolution versions. It means that they can improve the quality of their inputs by a factor of two. By preparing new databases containing new pairs of low and high resolution images of bigger sizes these algorithms can improve the quality by factors bigger than two. For factors bigger than two, however, there is no guarantee that they can provide true high resolution details [328]. Therefore, combining these two types of algorithms can lead to higher improvement factors [328]. Next chapter will introduce one of such algorithms that we have developed to work with the proposed system of this thesis.

The idea of using a similar hybrid super-resolution system has been used previously by [328]. They have used the concept of patch repetition for producing a high resolution output for a single low resolution input image. They produce a resolution pyramid for their input image. Then they look for the repeated patches inside the input image and across its resolution pyramid. Repeated patches within the same image scale are used in a reconstruction-based super resolution algorithm (they denote this algorithm as classical multi-image super resolution). While, in a recognition-based (they call it example based) super resolution algorithm repeated patches across the resolution pyramid are used.

6.6 Conclusion

This chapter gives a survey of the super-resolution algorithms in the literature. A schema for grouping these algorithms has been introduced. Then, the different techniques and their pros and cons are discussed. Meanwhile, a new group of hybrid super-resolution algorithms are mentioned which are more explained in the next chapter.

References

[24] R.W. Gerchberg, "*Super-resolution through error energy reduction*," Journal of Modern Optics, vol. 21, no. 9, pp. 709-720, 1974.

[25] P. De Santis and F. Gori, "*On an iterative method for super-resolution*," Journal of Modern Optics, vol. 22, no. 8, pp. 691-695, 1975.

[26] J.R. Fienup, "*Reconstruction of an object from the modulus of its Fourier transform*," Optics Letters, vol. 3, no. 1, pp. 27-29, 1978.

[27] M. Carlotto and V. Tom, "*Interpolation of two-dimensional surfaces using the Gerchberg algorithm*," SPIE, vol. 359, pp. 226-232, 1982.

[28] R. Franke, "*Scattered data interpolation: Tests of some methods*," Mathematics of Computation, vol. 38, no. 157, pp. 181-200, 1982.

[29] S. Geman and D. Geman, "*Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images*," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 6, no. 6, pp. 721-741, 1984.

[30] R. Tsai and T. Huang, "*Multiframe image restoration and registration*," In R. Y. Tsai and T. S. Huang, editors, Advances in Computer Vision and Image Processing, vol. 1, pp. 317-339, JAI Press Inc., 1984.

[31] J. Clark, M. Palmer and P. Laurence, "*A transformation method for the reconstruction of functions from non-uniformly spaced samples*," IEEE Transactions on Acoustics Speech and Signal Processing, 33(4), pp. 1151-1165, 1985.

[32] E. Mjolsness, "*Neural networks, pattern recognition, and fingerprint hallucination*," PhD thesis, California Institute of Technology, 1985.

[33] B. Frieden and H. Aumann, "*Image reconstruction from multiple 1-D scans using filtered localized projection*," Applied Optics, vol. 26, no. 17, pp. 3615-3621, 1987.

[34] S. Peleg, D. Keren and L. Schweitzer, "*Improving image resolution by using subpixel motion*," Pattern Recognition Letters, vol. 5. No. 3, pp. 223-226, 1987.

[35] D. Keren, S. Peleg and R. Brada, "*Image sequence enhancement using subpixel displacements*," Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 742-746, 1988.

[36] H. Stark and P. Oskoui, "*High resolution image recovery from image-plane arrays, using convex projections*," Journal of the Optical Society of America A, vol. 6, pp. 1715-1726, 1989.
[37] M. Irani and S. Peleg, "*Super-resolution from image sequences*," Proceedings of IEEE International Conference on Pattern Recognition, pp. 115-120, USA, 1990.
[38] A. Katsaggelos, "*A multiple input image restoration approach*," Journal of Visual Communication and Image Representation, vol. 1, no. 1, pp. 93-103, 1990.
[39] S. Kim and N. Bose, "*Reconstruction of 2-D bandlimited discrete signals from non-uniform samples*," Proceeding of the IEE, vol. 137, pp. 197-204, 1990.
[40] S.P. Luttrell, "*Bayesian autofocus/super-resolution theory*," Proceedings of IEE Colloquium on Role of Image Processing in Defence and Military Electronics, pp. 1-6, 1990.
[41] K. Aizawa, T. Komatsu and T. Saito, "*Acquisition of very high resolution images using stereo cameras*," Proceedings of Visual Communications and Image Processing, SPIE, vol. 1605, pp. 318-328, 1991.
[42] W.T. Freeman and E.H. Adelson, "*The design and use of steerable filters*," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 13, pp. 891-906, 1991.
[43] M. Irani and S. Peleg, "*Improving Resolution by Image Registration*," CVGIP: Graphical Models and Image Processing, vol. 53, pp. 231-239, 1991.
[44] A.K. Katsaggelos, Ed. "*Digital Image Restoration*," Heidelberg, Germany: Springer-Verlag. Springer. vol. 23, 1991.
[45] M. Irani and S. Peleg, "*Image sequence enhancement using multiple motions analysis*," Proceedings of International Conference on Computer Vision and Pattern Recognition, pp. 216-222, 1992.
[46] A. Schatzberg and A.J. Devaney, "*Super-resolution in diffraction tomography*," Inverse Problems, vol. 8, pp. 149-164, 1992.
[47] A. Tekalp, M. Ozkan and M. Sezan, "*High-resolution image reconstruction from lower-resolution image sequences and space-varying image restoration*," Proceedings of the IEEE International Conference on Acousics, Speech and Signal Processing, vol. III, pp. 169-172, USA, 1992.
[48] H. Ur and D. Gross, "*Improved resolution from sub-pixel shifted pictures*," CVGIP: Graphical Models and Image Processing, vol. 54, pp. 181-186, 1992.
[49] H.K. Aghajan and T. Kailath, "*Sensor array processing techniques for super-resolution multi-line-fitting and straight edge detection*," IEEE Transactions on Image Processing, vol. 2, no. 4, pp. 454-465, 1993.
[50] N. Bose, H. Kim and H. Valenzuela, "*Recursive implementation of total least squares algorithm for image reconstruction from noisy, undersampled multiframes*," Proceedings of the IEEE conference on Acoustics, Speech and Signal Processing, vol. 5, pp. 269-272, USA, 1993.
[51] M. Irani and S. Peleg, "*Motion analysis for image enhancement resolution, occlusion, and transparency*," Journal of Visual Communication and Image Representation, vol. 4, pp. 324-335, 1993.
[52] N. Bose, H. Kim and B. Zhou, "*Performance analysis of the TLS algorithm for image reconstruction from a sequence of undersampled noisy and blurred frames*," In Proceedings of the IEEE International Conference on Image Processing, vol. III, pp. 571-575, USA, 1994
[53] C. Davila, "*Efficient recursive total least squares algorithms for FIR adaptive filtering*," IEEE Transactions on Signal Processing, 42(2), pp. 268-280, 1994.

[54] L. Dumont, M. Fattouche and G. Morrison, "*Super-resolution of multipath channels in a spread spectrum location system*," Electronic Letters, vol. 30, no. 19, pp. 1583-1584, 1994.
[55] E. Fussfeld and Y.Y. Zeevi, "*Super-resolution estimation of edge images*," Proceedings of International Conference on Computer Vision & Image Processing, vol. 1, pp. 11-16, 1994.
[56] S. Mann, and R. Picard, "*Virtual bellows: Constructing high quality stills from video*," Proceedings of the IEEE conference on Image Processing, 1994.
[57] D.O. Walsh and P.A. Nielsen-Delaney, "*Direct method for super-resolution*," Journal of Optical Society of America A, vol. 11, pp. 572-579, 1994.
[58] Y. Murakami, A. Takahashi and S. Teashima, "*Magnetic super-resolution*," IEEE Transactions on Magnetic, vol. 31, no. 6, pp. 3215-3220, 1995.
[59] K. D. Sauer, S. Borman, and C. A. Bouman, ''*Parallel computation of sequential pixel updates in statistical tomographic reconstruction*,'' Proceedings of the IEEE International Conference on Image Processing, vol. 2, pp. 93-96, USA, 1995.
[60] H. Shekarforoush, M. Berthod, and J. Zerubia, "*3D super-resolution using generalized sampling expansion*," Proceedings of the IEEE International Conference on Image Processing, vol. 2, pp. 300-303, USA, 1995.
[61] *Z.D. Zh*u, Z.R. Ye, X.Q. Wu, J. Yin and Z.S. She, "*Super-resolution range-Doppler imaging*," IEE Proceedings on Radar, Sonar and Navigation, vol. 142, no. 1, pp. 25-32, 1995.
[62] B. Bascle, A. Blake, and A. Zisserman, "*Motion deblurring and super-resolution from an image sequence*," Proceedings of 4th European Conference on Computer Vision, pp. 312-320, UK, 1996.
[63] M.C. Chiang and T.E. Boult, "*Efficient image warping and super-resolution*," Proceedings of 3rd IEEE Workshop on Applications of Computer Vision, pp. 56-61, USA, 1996.
[64] M. Elad and A. Feuer, "*Super-resolution reconstruction of an image*," Proceedings of 19th IEEE Conference on Electrical and Electronics Engineers, pp. 391-394, Israel, 1996.
[65] M. Hanke, J.G. Nagy, "*Restoration of atmospherically blurred images by symmetric indefinite conjugate gradient techniques*," Inverse Problems, vol. 12, pp. 157-173, 1996.
[66] D. Kundur and D. Hatzinakos, "*Blind image deconvolution*," IEEE Magazine on Signal Processing, vol. 13, no. 3, pp. 43-64, 1996.
[67] C. Miller, B.R. Hunt, R.L. Kendrick and A.L. Duncan, "*Reconstruction and super-resolution of dilute aperture imagery*," Proceedings of International Conference on Image Processing, Switzerland, 1996.
[68] H. Shekarforoush, M. Berthod, J. Zerubia, and M. Werman, "*Sub-pixel bayesian estimation of albedo and height*," International Journal on Computer Vision, vol. 19, no. 3, pp. 289-300, 1996.
[69] D. Sheppard, B.R. Hunt and M.W. Marcellin, "*Super-resolution of imagery acquired through turbulent atmosphere*," Proceedings of 13th IEEE Conference on Signals, Systems and Computers, vol. 1, pp. 81-85, USA, 1996.
[70] B.C. Tom and A.K. Katsaggelos "*An iterative algorithm for improving the resolution of video sequences,*" Proceeding of SPIE Conference on Visual Communications and Image Processing, pp. 1430-1438, USA, 1996.
[71] B. Tom and A. Katsaggelos, "*Resolution enhancement of video sequences using motion compensation*," Proceedings of the IEEE International Confeence on Image Processing, vol. I, pp. 713-716, Switzerland, 1996.
[72] M. C. Chiang and T.E. Boult, "*Imaging-consistent super-resolution*," DARPA Image Understanding Workshop, 1997.

[73] M.C. Chiang and T.E. Boult, "*Local blur estimation and super-resolution*," Proceedings of the International Conference on Computer Vision and Pattern Recognition, pp. 821-826, Puerto Rico, 1997.
[74] M. Elad and A. Feuer, "*Restoration of a single super-resolution image from several blurred, noisy, and under-sampled measured images*," IEEE Transaction on Image Processing, vol. 6, no. 12, pp. 1646-1658, 1997.
[75] P.E. Eren, M.I. Sezan, and A.M. Tekalp, "*Robust object-based high-resolution image reconstruction from low-resolution video*," IEEE Transaction on Image Processing, vol. 6, no. 10, pp. 1446-1451, 1997.
[76] J.J. Green and B.R. Hunt, "*Super-resolution in a synthetic aperture imaging system*," Proceedings of International Conference on Image Processing, vol. 1, pp. 865-868, 1997.
[77] R.C. Hardie, K.J. Barnard and E.E. Armstrong, "*Joint MAP registration and high-resolution image estimation using a sequence of under sampled images*," IEEE Transaction on Image Processing, vol. 6, pp. 1621-1633, 1997.
[78] R.C. Hardie, K.J. Barnard, J.G. Bognar, E.E. Armstrong, and E.A. Watson, "*High-resolution image reconstruction from a sequence of rotated and translated frames and its application to an infrared imaging system*," Optical Engineering, vol. 37, no. 1, pp. 247-260, 1998.
[79] M.C. Hong, M.G. Kang, and A.K. Katsaggelos, "*A regularized multichannel restoration approach for globally optimal high resolution video sequence*," SPIE VCIP, vol. 3024, pp. 1306-1317, USA, 1997.
[80] M.C. Hong, M.G. Kang, and A.K. Katsaggelos, "*An iterative weighted regularized algorithm for improving the resolution of video sequences*," Proceeding of International Conference on Image Processing, vol. 2, pp. 474-477, 1997.
[81] A. Lorette, H. Shekarforoush, and J. Zerubia, "*Super-resolution with adaptive regularization*," Proceedings of the IEEE International Conference on Image Processing, vol. 1, pp. 169-172, USA, 1997.
[82] A.J. Patti, M. Sezan, and A.M. Tekalp, "*Robust methods for high quality stills from interlaced video in the presence of dominant motion*," IEEE Transactions on Circuits and Systems for Video Technology, vol. 7, no. 2, pp. 328-342, 1997.
[83] A.J. Patti, M.I. Sezan, and A.M. Tekalp, "*Superresolution video reconstruction with arbitrary sampling lattices and nonzero aperture time*," IEEE Transaction on Image Processing, vol. 6, no. 8, pp. 1064-1076, 1997.
[84] D. Calle and A. Montanvert, "*Super-resolution inducing of an image*," Proceedings of IEEE International Conference on Image Processing, vol. 3, pp. 232-235, 1998.
[85] D. Capel and A. Zisserman, "*Automated mosaicing with super-resolution zoom*," Proceedings of the International Conference on Computer Vision and Pattern Recognition, pp. 885-891, 1998.
[86] S. Borman and R.L. Stevenson, "*Spatial resolution enhancement of low-resolution image sequences A comprehensive review with directions for future research*," Laboratory of Image and Signal Analysis, University of Notre Dame, Technical Report, 1998.
[87] S. Borman and R.L. Stevenson, "*Super-resolution from image sequences-A review*," Proceedings of Midwest Symposium on Circuits and Systems, pp. 374-378, 1998.
[88] M.G. Kang, "*Generalized multichannel image deconvolution approach and its applications*," Optical Engineering, vol. 37, no. 11, pp. 2953-2964, 1998.
[89] D. Pastina, A. Farina, J. Gunning, and P. Lombardo, "*Two-dimensional super-resolution spectral analysis applied to SAR images*," IEE Proceedings of Radar, Sonar and Navigation, vol. 145, no.5, pp. 281-290, 1998.

[90] A. Patti and Y. Altunbasak, "*Artifact Reduction for POCS-based super-resolution with edge adaptive regularization and higher-order interpolants*," Proceedings of IEEE International Conference on Image Processing, pp. 217-221, USA, 1998.
[91] R. R. Schultz, L. Meng, and R. L. Stevenson, "*Subpixel motion estimation for super-resolution image sequence enhancement*," Journal of Visual Communication and Image Representation, vol. 9, no. 1, pp. 38-50, 1998.
[92] A. Zomet and S. Peleg, "*Applying super-resolution to panoramic mosaics*," Workshop on Applications of Computer Vision, 1998.
[93] S. Baker and T. Kanade, "*Hallucinating faces*," Technical Report CMU-RI-TR-99-32, The Robotics Institute, Carnegie Mellon University, 1999.
[94] S. Borman, M. Robertson and R. Stevenson, "*Block-matching sub-pixel motion estimation from noisy, under-sampled frames: an empirical performance evaluation*," SPIE Symposium on Visual Communication and Image Processing, USA, 1999.
[95] S. Borman and R.L. Stevenson, "*Simultaneous multi-frame MAP super-resolution video enhancement using spatiotemporal priors*," Proceedings of IEEE International Conference on Image Processing, pp. 469-473, Japan, 1999.
[96] F.M. Candocia and J.C. Principe, "*Super-resolution of images based on local correlations*," IEEE Transaction on Neural Networks, vol. 10, no. 2, pp. 372-380, 1999.
[97] M. Elad and A. Feuer, "*Super-resolution reconstruction of image sequences*," IEEE Transaction on Pattern Analysis on Machine Intelligence, vol. 21, no. 9, pp. 817-834, 1999.
[98] M. Elad and A. Feuer, "*Super-resolution restoration of an image sequence: adaptive filtering approach,"* IEEE Transaction on Image Processing, vol. 8, pp. 387-395, 1999.
[99] W.T. Freeman, and E. C. Pasztor, "*Learning to estimate scenes from images*," In: M. S. Kearns, S. A. Solla, and D. A. Cohn (eds.): Adv. Neural Information Processing Systems, Vol. 11. Cambridge, MA, 1999.
[100] W. Freeman and E. Pasztor, "*Markov networks for low-level vision*," Mitsubishi Electric Research Laboratory Technical Report TR99-08, 1999.
[101] B.R. Hunt, "*Imagery super-resolution: Emerging prospects*," Proceedings of SPIE on Applications of Digital Image Processing XIV, vol. 1567, pp. 600-608, USA, 1991.
[102] N. Nguyen, P. Milanfar, and G. Golub, "*Blind super-resolution with generalized cross-validation using gauss-type quadrature rules*," Proceedings of the 33rd Asilomar Conference on Signals, Systems, and Computers, 1999.
[103] M.C. Pan and A.H. Lettington, "*Efficient method for improving Poisson MAP super-resolution*," Electron Letters, vol. 35, pp. 803-805, 1999.
[104] M.A. Robertson, S. Borman, and R.L. Stevenson, "*Dynamic range improvement through multiple exposures*," Proceedings of the IEEE International Conference on Image Processing, pp. 159-163, Japan, 1999.
[105] H. Shekarforoush and R. Chellappa, "*Data-driven multi-channel super-resolution with application to video data*," Journal of Optical Society of America A, vol. 16, no. 3, pp. 481-492, 1999.
[106] S. Baker and T. Kanade, "*Limits on super-resolution and how to break them*," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 372-379, USA, 2000.
[107] S. Bhattacharjee and M.K. Sundareshan, "*Modeling and extrapolation of prior scene information for set-theoretic restoration and super-resolution of diffraction-limited images*," IEEE International Conference on Image Processing, Canada, 2000.

[108] N.K. Bose, S. Lertrattanapanich, and J. Koo, "*Advances in superresolution using L-curve*," Proceeding of International Symposium on Circuits and Systems, vol. 2, pp. 433-436, 2001.
[109] D. Capel, "*Super-Resolution enhancement of text image sequences*," Proceedings of the International Conference on Pattern Recognition, Spain, 2000
[110] S. Changyin and B. Zheng, "*Super-resolution algorithm for instantaneous ISAR imaging*," Electronics Letter, vol. 36, no. 3, pp. 253-255, 2000.
[111] M.C. Chiang and T.E. Boult, "*Efficient super-resolution via image warping*," Image Vision Computing, vol. 18, pp. 761-771, 2000.
[112] B. Cohen and I. Dinstein, "*Polyphase back-projection filtering for image resolution enhancement*," IEE Proceedings on Vision Image Signal Process, vol. 147, pp. 318-322, 2000.
[113] N.H. Dowlut and A. Manikas, "*A polynomial rooting approach to super-resolution array design*," IEEE Transaction on Signal Processing, vol. 48, pp. 1559-1569, 2000.
[114] S. Ebihara, M. Sato, and H. Niitsuma, "*Super-resolution of coherent targets by a directional borehole radar*," IEEE Transaction on Geoscience Remote Sensing, vol. 38, pp. 1725-1732, 2000.
[115] W.T. Freeman, E.C. Pasztor, and O.T. Carmichael, "*Learning low-level vision*," International Journal of Computer Vision, vol. 20, no. 1, pp. 25-47, 2000.
[116] T.F. Gee, T.P. Karnowski, and K.W. Tobin, "*Multiframe combination and blur deconvolution of video data*," Proceedings of SPIE on Image and Video Communications and Processing, vol. 3974, p. 788-795, 2000.
[117] N.X. Nguyen, "*Numerical algorithms for image super-resolution*," PhD thesis, Stanford University, 2000.
[118] V. Smelyanskiy, P. Cheeseman, D. Maluf, and R. Morris, "*Bayesian super-resolved surface reconstruction from images*," Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, USA, 2000.
[119] A. Zomet and S. Peleg, "*Efficient super-resolution and applications to mosaics*," Proceedings of IEEE International Conference on Pattern Recognition, pp. 579-583, Spain, 2000.
[120] S. Baker and T. Kanade, "*Super-resolution: reconstruction or recognition?*," Proceedings of IEEE EURASIP Workshop on Nonlinear Signal and Image Processing, USA, 2001.
[121] D.P. Capel, "*Image mosaicing and super-resolution*," PhD thesis, University of Oxford, 2001.
[122] D.P. Capel and A. Zisserman, "*Super-resolution from multiple views using learnt image models*," Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 627-634, USA, 2001.
[123] F. Dekeyser, P. Bouthemy, and P. Perez, "*A new algorithm for super-resolution from image sequences*," In: Computer Analysis of Images and Patterns, pp. 473-481, Germany, 2001.
[124] M. Elad and Y. Hel-Or, "*A fast super-resolution reconstruction algorithm for pure translational motion and common space-invariant blur*," IEEE Transaction on Image Processing, vol. 10, no. 8, pp. 1187-1193, 2001.
[125] A. Hertzmann, C. E. Jacobs, N. Oliver, B. Curless, and D. H. Salesin, "*Image analogies*," Proceedings of ACM Annual Conference Series SIGGRAPH, Computer Graphics, 2001.

[126] N. Jiang, R. Wu, and J. Li; "*Super-resolution feature extraction of moving targets*," IEEE Transaction on Aerospace and Electronic Systems, vol. 37, no. 3, pp. 781-793, 2001.

[127] I. Motoyoshi and S. Nishida, "*Temporal resolution of orientation-based texture segregation*," Vision Research, vol. 41, pp.2089-2105, 2001.

[128] H. Kim, J. H. Jang, and K. S. Hong, "*Edge-enhancing super-resolution using anisotropic diffusion*," Proceedings of IEEE International Conference on Image Processing, pp. 130-133, Greece, 2001.

[129] N. Nguyen, P. Milanfar, and G. Golub, "*A computationally efficient super-resolution image reconstruction algorithm*," IEEE Transaction on Image Processing, vol. 10, pp. 573-583, 2001.

[130] N. Nguyen, P. Milanfar, and G. Golub, "*Efficient generalized cross-validation with applications to parametric image restoration and resolution enhancement*," IEEE Transaction on Image Processing, vol. 10, pp. 1299-1308, 2001.

[131] A.J. Patti and Y. Altunbasak, "*Artifact reduction for set theoretic super-resolution image reconstruction with edge adaptive constraints and higher-order interpolants*," IEEE Transaction on Image Processing, vol. 10, no. 1, pp. 179-186, 2001.

[132] H. Szu and I. Kopriva, "*Artificial neural networks for noisy image super-resolution*," Optics Communications, vol. 198, pp. 71-81, 2001.

[133] D. Rajan and S. Chaudhuri, "*Generalized interpolation and its applications in super-resolution imaging*," Image and Vision Computing, vol. 19, pp. 957-969, 2001.

[134] D. Rajan and S. Chaudhuri, "*Generation of super-resolution images from blurred observations using Markov random fields*," Proceedings of IEEE International Conference on Acoustics, Speech, Signal Processing, vol. 3, pp. 1837-1840, USA, 2001.

[135] C.A. Segall, A.K. Katsaggelos, R. Molina, and J. Mateos, "*Super-resolution from compressed video*," In Super-Resolution Imaging, S. Chaudhuri, Ed. Boston, MA: Kluwer, pp. 211-242, 2001.

[136] B.C. Tom, N.P. Galatsanos, and A.K. Katsaggelos, "*Reconstruction of a high resolution image from multiple low resolution images*," In Super-Resolution Imaging, S. Chaudhuri, Ed. Norwell, MA: Kluwer, ch. 4, pp. 73-105, 2001.

[137] A. Zomet and S. Peleg, "*Super-resolution from multiple images having arbitrary mutual motion*," In: Chaudhuri S, editor. Super-resolution imaging. Norwell (MA): Kluwer Academic, pp. 195-209, 2001.

[138] A. Zomet, A. Rav-Acha, and S. Peleg, "*Robust super-resolution*," Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 645-650, USA, 2001.

[139] Y. Altunbasak, A.J. Patti, and R.M. Mersereau, "*Super-resolution still and video reconstruction from mpeg-coded video*," IEEE Transactions on Circuits and Systems for Video Technology, vol. 12, pp. 217-226, 2002.

[140] T.E. Boult, M.C. Chiang, and R.J. Micheals, "*Super-resolution via image warping*," In Chaudhuri S, editor. Super-resolution imaging. Norwell (MA): Kluwer Academic, pp. 131-169, 2002.

[141] W.T. Freeman, T.R. Jones, and E.C. Pasztor, "*Example-Based Super-Resolution*," IEEE Computer Graphics and Applications, vol. 22 no.2, pp.56-65, 2002.

[142] B.K. Gunturk, Y. Altunbasak, and R.M. Mersereau, "*Multiframe resolution-enhancement methods for compressed video*," IEEE Signal Processing Letters, vol. 9, pp. 170-174, 2002.

[143] H.V. Le and Guna Seetharaman, "*A super-resolution imaging method based on dense subpixel-accurate motion fields*," Proceedings of 3rd International Workshop on Digital and Computational Video, pp. 35-42, 2002.
[144] U. Nickel, "*Spotlight MUSIC: Super-resolution with subarrays with low calibration effort*," IEE Proceedings on Radar, Sonar, Navigation, vol. 149, no. 4, pp. 166-173, 2002.
[145] D. Rajan and S. Chaudhuri, "*An MRF-based approach to generation of super-resolution images from blurred observations*," Journal of Mathematical Imaging and Vision, vol. 16, no.1, pp. 5-15, 2002.
[146] D. Rajan and S. Chaudhuri, "*Data fusion techniques for super-resolution imaging*," Information Fusion, vol. 3, no. 1, pp. 25-38, 2002.
[147] A. Storkey, "*Dynamic structure super-resolution*," Advances in Neural Information Processing Systems, vol. 16, pp. 1295-1302, 2002.
[148] M.E. Tipping and C.M. Bishop, "*Bayesian image super-resolution*," Advances in Neural Information Processing Systems, vol. 15, pp. 1303-1310, 2002.
[149] W. Zhao, H. Sawhney, M. Hansen, and S. Samarasekera, "*Super-fusion: a super-resolution method based on fusion*," Proceedings of IEEE International Conference on Pattern Recognition, vol. 2, pp. 269-272, Canada, 2002.
[150] A. Zomet and S. Peleg, "*Multi-sensor super-resolution*," Proceedings of 6th IEEE Workshop on Applications of Computer Vision, pp. 27-31, USA, 2002.
[151] J. Abad, M. Vega, R. Molina, and A.K. Katsaggelos, "*Parameter estimation in super-resolution image reconstruction problems*," IEEE International Conference on Acoustic, Speech and Signal Processing, vol. 3, pp. 709-712, Hong Kong, 2003.
[152] M. Bertero and P. Boccacci, "*Super-resolution in computational imaging*," Micron, vol. 34, pp. 265-273, 2003.
[153] C. Bishop, A. Blake, and B. Marthi, "*Super-resolution enhancement of video*," Proceedings of Artificial Intelligence and Statistics, 2003.
[154] S. Borman and R. L. Stevenson, "*Image resampling and constraint formulation for multi-frame super-resolution restoration*," Proceedings of SPIE on Computational Imaging, vol. 5016, pp. 208-219, USA, 2003.
[155] G. Canel, A.M. Tekalp, and W. Heinzelman, "*Super-resolution recovery for multi-camera surveillance imaging*," Proceedings of IEEE International Conference on Multimedia and Expo, pp.109-112, USA, 2003.
[156] D. Capel and A. Zisserman, "*Computer vision applied to super-resolution*," IEEE Signal Processing Magazine, vol. 20, no. 3, pp. 75-86, 2003.
[157] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "*Robust shift and add approach to super-resolution*," Proceedings of SPIE Conference on Applications of Digital Signal and Image Processing, pp.121-130, USA, 2003.
[158] A. Fitzgibbon, Y. Wexler, and A. Zisserman, "*Image-based rendering using image-based priors*," Proceedings of 9th IEEE International Conference on Computer Vision, vol. 2, pp.1176-1183, China, 2003.
[159] N. Goldberg, A. Feuer, and G. C. Goodwin, "*Super-resolution reconstruction using spatio-temporal filtering*," Journal of Visual Communication and Image Representation, vol. 14, no. 4, pp. 508-525, 2003.
[160] B.K. Gunturk, A.U. Batur, Y. Altunbasak, M.H. Hayes, and R.M. Mersereau, "*Eigenface based super-resolution for face recognition*," IEEE Transaction on Image Processing, pp. 597-606, 2003.

[161] Z. Jiang, T.T. Wong, and H. Bao, "*Practical super-resolution from dynamic video sequences*," Proceedings of International Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 549-554, Canada, 2003.

[162] M. V. Joshi and S. Chaudhuri, "*A learning-based method for image super-resolution from zoomed observations*," Proceedings of 5th International Conference on Advances in Pattern Recognition, pp. 179-182, India, 2003.

[163] C. Miravet and F.B. Rodriguez, "*A hybrid MLP-PNN architecture for fast image super-resolution*," International Conference on Neural Information Processing, pp. 417-424, Turkey, 2003.

[164] M.K. Ng and N.K. Bose, "*Mathematical analysis of super-resolution methodology*," IEEE Signal Processing Magazine, vol. 20, no. 3, pp. 62-74, 2003.

[165] S.C. Park, M. K. Park, and M. G. Kang, "*Super-resolution image reconstruction: a technical overview*," IEEE Signal Processing Magazine, vol. 20, no. 3, pp. 21-36, 2003.

[166] L. Pickup, S. Roberts, and A. Zisserman, "*A sampled texture prior for image super-resolution*," Proceeding of 16th International conference on Advances in Neural Information Processing Systems, 2004.

[167] D. Rajan, S. Chaudhuri, and M.V. Joshi, "*Multi-objective super-resolution: concepts and examples*," IEEE Signal Processing Magazine, vol. 20, no. 3, pp. 49-61, 2003.

[168] D. Rajan and S. Chaudhuri, "*Simultaneous estimation of super-resolved scene and depth map from low resolution observations*," IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 25, pp. 1102-1117, 2003.

[169] E. Salari, S. Zhang, "*Integrated recurrent neural network for image resolution enhancement from multiple image frames*," IEE Vision, Image, and Signal Processing, vol. 150, no. 5, pp. 299-305, 2003.

[170] C.A. Segall, R. Molina, and A.K. Katsaggelos, "*High-resolution images from low-resolution compressed video*," IEEE Signal Processing Magazine, vol. 20, no. 3, pp. 37-48, 2003.

[171] M. F. Tappen, B. C. Russell, and W. T. Freeman, "*Exploiting the sparse derivative prior for super-resolution and image demosaicing*," IEEE Workshop Statistical and Computational Theories of Vision, 2003.

[172] P. Vandewalle, S. E. Susstrunk, and M. Vetterli, "*Super-resolution images reconstructed from aliased images*," Proceedings of SPIE Conference on Visual Communications and Image Processing, vol. 5150, pp. 1398-1405, Switzerland, 2003.

[173] X. Wang and X. Tang, "*Face hallucination and recognition*," Proceedings of 4th International Conference on Audio- and Video-based Personal Authentication, IAPR, pp. 486-494U.K., 2003.

[174] L. Yin and M. T. Yourst, "*3D face recognition based on highresolution 3D face modeling from frontal and profile views*," Proceedings of ACMSIGMM workshop on Biometrics methods and applications, pages 1-8, 2003.

[175] G. Zweig, "*Super-resolution Fourier transform by optimization, and ISAR imaging*," IEE Proceedings on Radar, Sonar and Navigation pp. 247-252, 2003.

[176] A. Almansa, S. Durand, and B. Roug´e, "*Measuring and improving image resolution by adaptation of the reciprocal cell*," Journal of Mathematical Imaging and Vision, vol. 21, pp. 235-279, 2004.

[177] I. Begin and F. P. Ferrie: "*Blind Super-Resolution Using a Learning-Based Approach*," Proceedings of IEEE International Conference on Pattern Recognition, pp. 85-89, UK, 2004.

[178] M. Ben-Ezra, and S. Nayar, "*Jitter camera: High resolution video from a low resolution detector*," Proceeding of IEEE International Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 135-142, USA, 2004.

[179] S. Borman, "*Topics in multiframe super-resolution restoration*," PhD thesis, University of Notre Dame, 2004.

[180] S. Borman and R. Stevenson, "*Linear models for multi-frame super-resolution restoration under non-affine registration and spatially varying psf*," SPIE Electronic Imaging, 2004.

[181] N.K. Bose, S. Lertrattanapanich, and M.B. Chappali, "*Super-resolution with second generation wavelets*," Signal Processing Image Communication, vol. 19, pp. 387-391, 2004.

[182] H. Chang, D.Y. Yeung, and Y. Xiong, "*Super-resolution through neighbor embedding*," Proceeding of IEEE International Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 275-282, USA, 2004.

[183] J. Cui, Y. Wang, J. Huang, T. Tan, and Z. Sun, "*An iris image synthesis method based on PCA and super-resolution*," Proceedings of IEEE International Conference on Pattern Recognition, vol. 4, pp. 471-474, UK, 2004.

[184] G. Dedeoglu, T. Kanade, and J. August, "*High-zoom video hallucination by exploiting spatio-temporal regularities*," Proceeding of IEEE International Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 151-158, USA, 2004.

[185] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "*Advances and challenges in super-resolution*," International Journal of Imaging Systems and Technology, vol. 14, no. 2, pp. 47-57, 2004.

[186] S Farsiu, D. Robinson, M. Elad, P. Milanfar, "*Dynamic demosaicing and color Super-Resolution video sequences*," Proceedings of SPIE Conference on Image Reconstruction from Incomplete Data, 2004.

[187] S. Farsiu, D. Robinson, M. Elad, P. Milanfar. 2004, "*Fast and robust multi-frame super-resolution*," IEEE Transaction on Image Processing, vol. 13, no. 10, pp. 1327-1344, 2004.

[188] S. Farsiu, M. Elad, P. Milanfar, "*Multi-frame demosaicing and super-resolution from under-sampled color images*," Proceeding of SPIE Symposium on Electronic Imaging, pp. 222-233, 2004.

[189] B.K. Gunturk, Y. Altunbasak, and R.M. Mersereau, "*Super-resolution reconstruction of compressed video using transform-domain statistics*," IEEE Transaction on Image Processing, vol. 13, no. 1, pp. 33-43, 2004.

[190] C.V. Jiji, M.V. Joshi, and S. Chaudhuri, "*Single-frame image super-resolution using learned wavelet coefficients*," International Journal of Imaging Systems and Technology, vol. 14, no. 3, pp. 105-112, 2004.

[191] Y. Li and X. Lin, "*An improved two-step approach to hallucinating faces*" Proceedings of 3rd International Conference on Image and Graphics, pp. 298-301, Hong Kong, 2004.

[192] Z. Lin and H.Y. Shum, "*Fundamental limits of reconstruction-based super-resolution algorithms under local translation*," IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 26, pp. 83-97, 2004.

[193] C.A. Segall , A.K. Katsaggelos , R. Molina and J. Mateos "*Bayesian resolution enhancement of compressed video*", IEEE Transaction on Image Processing, vol. 13, pp. 898, 2004.

[194] S. Villena, J. Abad, R. Molina, and A. K. Katsaggelos, "*Estimation of high resolution images and registration parameters from low resolution observations*," Iberoamerican Congress on Pattern Recognition, pp. 509-516, Mexico, 2004.

[195] L. Yin, J. Loi, and W. Xiong, "*Facial expression representation and recognition based on texture augmentation and topographic masking*," ACM Multimedia, 2004.

[196] Z. Wang and F. Qi, "*On ambiguities in super-resolution modeling*," IEEE Signal Processing Letters, vol. 11, no.8, pp. 678-681, 2004.

[197] T. Akgun, Y. Altunbasak, and R. M. Mersereau, "*Super-resolution reconstruction of hyperspectral images*," IEEE Transaction on Image Processing, vol. 14, pp. 1860-1875, 2005.

[198] D. Barreto, L. D. Alvarez, J. Abad, "*Motion estimation techniques in super-resolution image reconstruction: a performance evaluation*," Proceedings of Virtual Observatory: Plate Content Digitalization, Archive Mining and Image Sequence Processing, Bulgaria, 2005.

[199] M. Ben-Ezra, A. Zomet, and S.K. Nayar, "*Video super-resolution using controlled subpixel detector shifts*," IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 27, no. 6, pp. 977-987, 2005.

[200] M. Chappalli and N. Bose, "*Simultaneous noise filtering and super-resolution with second-generation wavelets*," Signal Processing Letters, vol. 12, pp. 772-775, 2005.

[201] K. Donaldson and D. K. Myers, "*Bayesian super-resolution of text in video with a text-specific bimodal prior*," International Journal on Document Analysis and Recognition, vol. 7, no. 2, pp. 159-167, 2005.

[202] H. He and L. P. Kondi, "*A regularization framework for joint blur estimation and super-resolution of video sequences*," Proceedings of IEEE International Conference on Image Processing, vol. 3, pp. 11-14, Italy, 2005.

[203] S. Farsiu, "*A fast and robust framework for image fusion and enhancemen*t," PhD thesis, University of California, Santa Cruz, 2005.

[204] M. Gevrekci, B.K. Gunturk "*Image acquisition modeling for super-resolution reconstruction*," IEEE International Conference on Image Processing, vol.2, pp.1058-1061, Italy, 2005.

[205] M.D. Gupta, S. Rajaram, N. Petrovic, and T.S. Huang, "*Non-parametric image super-resolution using multiple images*," Proceedings of IEEE International Conference on Image Processing, Italy, 2005.

[206] K. Jia and S. Gong, "*Multi-modal tensor face for simultaneous super-resolution and recognition*," Proceeding of International Conference on Computer Vision, vol. 2, pp. 1683-1690, China, 2005.

[207] T. Kasetkasem, M.K. Arora, and P.K. Varshney, "*Super-resolution land cover mapping using a Markov random field based approach*," Remote Sensing of Environment, vol. 96, 302-314, 2005.

[208] F. Lin, C. Fookes, V. Chandran, and S. Sridharan, "*Investigation into optical flow super-resolution for surveillance applications*," Proceedings of APRS Workshop on Digital Image Computing, February 2005, pp. 73–78.

[209] W. Liu, D. Lin, and X. Tang, "*Hallucinating faces: Tensorpatch super-resolution and coupled residue compensation*," Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 478-484, USA, 2005.

[210] C. Mancas-Thillou and M. Mirmehdi, "*Super-resolution text using the teager filter*," Proceedings of 1st International Workshop on Camera-Based Document Analysis and Recognition, pp. 10-16, Korea, 2005.

[211] C. Miravet and F. B. Rodrıguez, "*Accurate and robust image super-resolution by neural processing of local image representations*," Proceedings of International Conference on Artificial Neural Networks, vol. 1, pp. 499-505, Poland, 2005.

[212] M.K. Ng and A.C. Yau, "*Super-resolution image restoration from blurred low-resolution images*," Journal of Mathematical Imaging and Vision, vol. 23, no. 3, pp. 367-378, 2005.

[213] T.Q. Pham, L.J. van Vliet, and K. Schutte, "*Robust fusion of irregularly sampled data using adaptive normalized convolution*," EURASIP Journal on Advances in Signal Processing, 12 pages, Article ID 83268, 2005.

[214] C. Papathanassiou, and M. Petrou, "*Super-resolution: an overview*," Proceedings of International Symposium on Geoscience and Remote Sensing, vol. 8, pp. 5655-5658, Korea, 2005.

[215] J. Park, Y. Kwon, and J. H. Kim, "*An example-based prior model for text image super-resolution*," Proceedings of IEEE 8th International Conference on Document Analysis and Recognition, vol.1, pp. 374-378, Korea, 2005.

[216] C. Rubert, L. Fonseca, and L. Velho, "*Learning based super-resolution using YUV model for remote sensing images*," Proceedings of WTDCGPI, 2005.

[217] E. Shechtman, Y. Caspi, and M. Irani, "*Space-time super-resolution*," IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 27, no. 4, pp. 531-545, 2005.

[218] C. Su, L. Huang, "*Facial expression hallucination*," Proceedings of 7th IEEE Workshop on Application of Computer Vision, vol. 1, pp. 93-98, USA, 2005.

[219] C. Su, Y. Zhuang, L. Huang, and F. Wu, "*Steerable pyramid based face hallucination*," Pattern Recognition, vol. 38, pp. 813-824, 2005.

[220] K. Su, Q. Tian, Q. Que, N. Sebe, and J. Ma, "*Neighborhood issue in single-frame image super-resolution*," Proceedings of IEEE International Conference on Multimedia and Expo, The Netherlands, 2005.

[221] Z. Wang and F. Qi, "*Analysis of multiframe super-resolution reconstruction for image anti-aliasing and deblurring*," Image and Vision Computing, vol. 23, pp. 393-404, 2005.

[222] X. Wang and X. Tang, "*Hallucinating face by eigentransformation*," IEEE Transaction on Systems, Man and Cybernetics, vol. 35, no. 3, pp. 425-434, 2005.

[223] Q. Wang, X. Tang, and H. Shum, "*Patch based blind image super-resolution*," Proceedings of 10th International Conference on Computer Vision, vol. 1, pp. 709-716, China, 2005.

[224] X. Xu and R. Narayanan, "*Enhanced resolution in SAR/ISAR imaging using iterative sidelobe apodization*,' IEEE Transactions on Image Processing, vol. 14, no. 4, pp. 537-547, 2005.

[225] D. Zhang, H. Li, and M. Du, "*Fast MAP-based multiframe super-resolution image reconstruction*," Image and Vision Computing, no. 23, pp. 671-679, 2005.

[226] M.V.W. Zibetti and J. Mayer, "*Simultaneous super-resolution for video sequences*," Proceedings of IEEE International Conference on Image Processing, vol. 1, pp. 877-880, Italy, 2005.

[227] J. Yu, B. Bhanu, "*Super-resolution restoration of facial images in video*," Proceedings of IEEE 18th International Conference on Pattern Recognition, vol. 4, pp. 342-345, Hong Kong, 2006.

[228] N.K. Bose, M.K. Ng, and A.C. Yau, "*A fast algorithm for image super-resolution from blurred observations*," EURASIP Journal on Advances in Signal Processing, 14 pages, Article ID 35726, 2006.

[229] G.M. Callicó, R.P. Llopis, S. López, J.F. Lopez, A. Nunez, R. Sethuraman, R. Sarmiento, "*Low-cost super-resolution algorithms implementation over a HW/SW video compression platform*," EURASIP Journal on Advances in Signal Processing, 29 pages, Article ID 84614, 2006.

[230] B. Choi and J.B Ra, "*Region-based super-resolution using multiple blurred and noisy under-sampled images*," Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, Toulouse, vol. 2, pp. 609-612, France, 2006.

[231] J. Chung, E. Haber, and J. Nagy, "*Numerical methods for coupled super-resolution*," Inverse Problems, vol. 22, no. 4, pp. 1261–1272, 2006.

[232] S. Farsiu, M. Elad, and P. Milanfar, "*A practical approach to super-resolution*," Proceedings of SPIE: Visual Communications and Image Processing, USA, 2006.

[233] S. Farsiu, M. Elad, and P. Milanfar, "*Multiframe demosaicing and super-resolution of color images*," IEEE Transactions on Image Processing, vol. 15, no. 1, pp. 141-159, 2006.

[234] S. Farsiu, M. Elad, P. Milanfar, "*Video-to-video dynamic super-resolution for grayscale and color sequence*s," EURASIP Journal of Applied Signal Processing, 15 pages, Article ID 61859, 2006.

[235] J.P. Fouque, J. Garnier and K. S lna, "*Time reversal super-resolution in randomly layered media*," Wave Motion, vol. 42, pp. 238-260, 2006.

[236] B.K. Gunturk and M. Gevrekci, "*High-resolution image reconstruction from multiple differently exposed images*," Signal Processing Letters, vol. 13, no. 4, pp. 197-200, 2006.

[237] H. He and L. P. Kondi, "*An image super-resolution algorithm for different error levels per frame*," IEEE Transaction on Image Processing, vol. 15, no.3, pp. 592-603, 2006.

[238] F. Humblot, and A. Muhammad-Djafari, "*Super-resolution using hidden Markov model and Bayesian detection estimation framework*," EURASIP Journal on Advances in Signal Processing, 16 pages, Article ID 36971, 2006.

[239] K. Jia and S. Gong, "*Hallucinating multiple occluded face images of different resolutions*," Pattern Recognition Letters, vol. 27, no. 15, pp.1768-1775, 2006.

[240] C.V. Jiji and S. Chaudhuri, "*Single-frame image super-resolution through contourlet learning*," EURASIP Journal on Advances in Signal Processing, 11 pages, Article ID 73767, 2006.

[241] M.V. Joshi and S. Chaudhuri, "*Simultaneous estimation of super-resolved depth map and intensity field using photometric cue*," Computer Vision and Image Understanding, vol. 101, no. 1, pp. 31-44, 2006.

[242] M. Lerotic and G.Z. Yang, "*The use of super-resolution in robotic assisted minimally invasive surgery*," Medical Image Computing and Computer-Assisted Intervention, pp. 462-469, 2006.

[243] X.Li, "*Super-Resolution for Synthetic Zooming*," EURASIP Journal on Advances in Signal Processing, 12 pages, Article ID 58195, 2006.

[244] R. Molina, J. Mateos, and A.K. Katsaggelos, "*Blind deconvolution using a variational approach to parameter, image, and blur estimation*," IEEE Transaction on Image Processing, vol. 15, no. 12, pp. 3715-3727, 2006.

[245] U. Mudenagudi, R. Singla, P. Kalra, and S. Banerjee, "*Super-resolution using graph-cut*," Proceedings of 7th Asian Conference on Computer Vision, pp. 385-394, India, 2006.

[246] L.C. Pickup, D.P. Capel, S.J. Roberts, and A. Zisserman, "*Bayesian image super-resolution, continued*," Neural Information Processing Systems, vol. 19, pp. 1089-1096, 2006.

[247] M. Pan, "*Improving a Single Down-Sampled Image Using Probability-Filtering-Based Interpolation and Improved Poisson-Maximum A Posteriori Super-Resolution*," EURASIP Journal on Advances in Signal Processing, 9 pages, Article ID 97492, 2006.

[248] G. Pan, S. Han, Z. Wu, and Y. Wang, "*Super-resolution of 3d face*," Proceedings of 9th European Conference on Computer Vision, vol. 3952, pp. 389-401, Austria, 2006.

[249] V. Patanavijit and S. Jitapunkul, "*An iterative super-resolution reconstruction of image sequences using affine block-based registration*," ACM International Symposium on Multimedia Over Wireless, Canada, 2006.

[250] V. Patanavijit, S. Jitapunkul, "*An iterative super-resolution reconstruction of image sequences using fast affine block-based registration with BTV regularization*," Proceedings of IEEE Asia Pacific Conference on Circuits and Systems, pp. 1717-1720, 2006.

[251] S. Rajaram, M.D. Gupta, N. Petrovic, and T. S. Huang, "*Learning based nonparametric image super-resolution*," EURASIP Journal on Advances in Signal Processing, 11 pages, Article ID 51306, 2006.

[252] D. Robinson, P. Milanfar, "*Statistical performance analysis of super-resolution*", IEEE Transaction on Image Processing, vol. 15, no. 6, pp. 1413-1428, 2006.

[253] G. Rochefort, F. Champagnat, G.L. Besnerais, and J.F. Giovannelli, "*An improved observation model for super-resolution under affine motion*," IEEE Transaction on Image Processing, vol. 15, no. 11, pp. 3325-3337, 2006.

[254] F. Sroubek and J. Flusser, "*Resolution enhancement via probabilistic deconvolution of multiple degraded images*," Pattern Recognition Letters, vol. 27, pp. 287-293, 2006.

[255] T. A. Stephenson and T. Chen, "*Adaptive markov random fields for example-based super-resolution of faces*," EURASIP Journal on Advances in Signal Processing, 11 pages, Article ID 31062, 2006.

[256] R. Szeliski, "*Image alignment and stitching: A tutorial*," Technical Report MSR-TR-2004-92, Microsoft Research, 2006.

[257] M. Trimeche, R.C. Bilcu, and J. Yrjanainen, "*Adaptive outlier rejection in image super-resolution*," EURASIP Journal on Advances in Signal Processing, 12 pages, Article ID 38052, 2006.

[258] P. Vandewalle, "*Super-resolution from unregistered aliased images*," PhD thesis, Ecole Polytechnique Federale de Lauasnne, 2006.

[259] P. Vandewalle, S. Susstrunk, and M. Vetterli, "*A frequency domain approach to registration of aliased images with application to super-resolution*," EURASIP Journal on Advances in Signal Processing, 14 pages, Article ID 71459, 2006.

[260] J.D. Van Ouwerkerk, "*Image super-resolution survey*," Image and Video Computing, vol. 24, no. 10, pp. 1039-1052, 2006.

[261] A.W.M. Van Eekeren, K. Schutte, J. Dijk, D. de Lange, and L. van Vliet, "*Super-resolution on moving objects and background*," Proceedings of the IEEE International Conference on Image Processing, pp. 2709-2712, USA, 2006.

[262] M. Vega, R. Molina, and A. Katsaggelos, "*A Bayesian super-resolution approach to demosaicing of blurred images*," EURASIP Journal on Advances in Signal Processing, 12 pages, Article ID 25072, 2006.

[263] C. Wang, P. Xue, and W. Lin, "*Improved super-resolution reconstruction from video*," IEEE Transactions on Circuits and Systems for Video Technology, vol. 16, no. 11, pp. 1411-1422, 2006.

[264] S. Zhang, "*Application of super-resolution image reconstruction to digital holography*," EURASIP Journal on Advances in Signal Processing, 7 pages, Article ID 90358, 2006.

[265] M.V.W. Zibetti, J. Mayer, "*Outlier robust and edge-preserving simultaneous super-resolution*," Proceedings of IEEE International Conference on Image Processing, vol. 1,pp. 1741-17441, USA, 2006.

[266] A. Ashok and M. A. Neifeld, "*Pseudorandom phase masks for super-resolution imaging from subpixel shifting*," Applied Optics, vol. 46, no. 12, pp. 2256-2268, 2007.

[267] L. Baboulaz and P.L. Dragotti, "*Local feature extraction for image super-resolution*," Proceedings of IEEE International Conference on Image Processing, pp. 401, USA, 2007.

[268] V. Bannore and L. Swierkowski, "*An iterative approach to image super-resolution*," Proceedings of International Federation for Information Processing, vol. 228, pp. 510-517, 2007.

[269] N. Damera-Venkata and N. L. Chang, "*Realizing superresolution with superimposed projection*," Proceedings of IEEE International Workshop on Projector-Camera Systems (ProCams), USA, 2007.

[270] D. Datsenko and M. Elad, "*Example-based single document image super-resolution: A global map approach with outlier rejection*," Journal of Multidimensional Systems and Signal Processing, no. 2, pp. 103-121, 2007.

[271] R.C. Hardie, "*A fast image super-resolution algorithm using an adaptive Wiener filter*," IEEE Transaction on Image Processing, vol. 16, pp. 2953-2964, 2007.

[272] E.Y.T. Ho and A.E. Todd-Pokropek, "*Further improvement of super-resolution reconstruction*," ICSIE, World Congress of Engineering, UK, 2007.

[273] M. Elad and D. Datsenko, "*Example-based regularization deployed to super-resolution reconstruction of a single image*," The Computer Journal, vol. 18, no. 2, pp. 103-121, 2007.

[274] R. Fattal, "*Image upsampling via imposed edge statistics*," ACM Special Interest Group on Computer Graphics and Interactive Techniques, vol. 26, no. 3, article 95, 8 pages, USA, 2007.

[275] R. Fransens, C. Strecha, and L. V. Gool, "*Optical flow based super-resolution: A probabilistic approach*," Computer Vision and Image Understanding, vol. 106, no. 1, pp. 106-115, 2007.

[276] S. Keller, F. Lauze, M. Nielsen, "*Motion compensated video super resolution*," In: Sgallari, F., Murli, A., Paragios, N. (eds.) SSVM LNCS, vol. 4485, pp. 801–812. 2007.

[277] W. Liu, X. Tang, and J. Liu, "*Bayesian tensor inference for sketchbased facial photo hallucination*," Proceedings of 20th International Joint Conference on Artificial Intelligence, pp. 2141-2146, India, 2007.

[278] C. Liu, H.Y. Shum, and W.T. Freeman "*Face hallucination: Theory and practice*," International Journal of Computer Vision, vol. 75, no. 1, pp. 115-134, 2007.

[279] S. Lui, J. Wu, H. Mao, and J. J. Lien, "*Learning-based super-resolution system using single facial image and multi-resolution wavelet synthesis*," Proceedings of Asian Conference on Computer Vision, vol. 4884, pp. 96-105, Japan, 2007.

[280] C. Miravet, and F.B. Rodriguez, "*A two step neural network based algorithm for fast image super-resolution*," Image and Vision Computing, vol. 25, pp. 1499-1473, 2007.

[281] B. Narayanan, R.C. Hardie, K.E. Barner, and M. Shao, "*A computationally efficient super-resolution algorithm for video processing using partition filters*," IEEE Transaction Circuits Systems Video Technology, vol. 17, no.5, pp. 621-634, 2007.

[282] M.K. Ng, H. Shen, E.Y. Lam, and L. Zhang, "*A total variation regularization based super-resolution reconstruction algorithm for digital video*," EURASIP Journal on Advances in Signal Processing, 16 pages, Article ID 74585, 2007.

[283] V. Patanavijit and S. Jitapunkul, "*A Lorentzian stochastic estimation for a robust iterative multiframe super-resolution reconstruction with Lorentzian-Tikhonov regularization*," EURASIP Journal on Advances in Signal Processing, Article ID 34821, 21 pages, 2007.

[284] S.W. Park, M. Savvides, "*Breaking the limitation of manifold analysis for super-resolution of facial images*," IEEE International Conference on Acoustics, Speech and Signal Processing, vol. 1, pp. 573–576, 2007.

[285] L.C. Pickup, "*Machine learning in multi-frame image super-resolution*," PhD thesis, University of Oxford, 2007.

[286] L.C. Pickup, D.P. Capel, S.J. Roberts, and A. Zisserman, "*Bayesian methods for image super-resolution*," The Computer Journal, vol. 52, pp. 101-113, 2007.

[287] L.C. Pickup, D.P. Capel, S. J. Roberts, and A. Zisserman, "*Overcoming registration uncertainty in image super-resolution: Maximize or marginalize?*," EURASIP Journal on Advances in Signal Processing, Article ID 23565, 14 pages, 2007.

[288] H.F. Shen, L.P. Zhang, B. Huang, and P.X. Li, "*A MAP approach for joint motion estimation, segmentation, and super-resolution*," IEEE Transaction on Image Processing, vol. 16, no. 2, pp. 479-490, 2007.

[289] D. Shengyang, H. Mei, W. Ying, and G. Yihong, "*Bilateral back-projection for single image super resolution*," Proceedings of IEEE International Conference on Multimedia and Expo, pp.1039-1042, 2007.

[290] F. Sroubek, G. Cristobal, and J. Flusser, "*A unified approach to super-resolution and multichannel blind deconvolution*," IEEE Transaction on Image Processing, vol. 16, pp. 2322-2332, 2007.

[291] A.W.M. van Eekeren, K. Schutte, O.R. Oudegeest and L.J. van Vliet, "*Performance evaluation of super-resolution reconstruction methods on real-world data*," EURASIP Journal on Advances in Signal Processing, 2007, Article ID 43953, 11 pages, 2007.

[292] F. Wheeler, X. Liu, and P. Tu, "*Multi-Frame Super-Resolution for Face Recognition*," Proceeding of IEEE Conference on Biometrics: Theory, Applications and Systems, pp. 27-29, USA, 2007.

[293] H. Yan, J. Liu, J. Sun, and X. Sun, "*ICA based super-resolution face hallucination and recognition*," 4th International Symposium on Neural Networks, vol. 2, pp. 1065-1071, 2007.

[294] S.T. Zhang and Y.H. Lu, "*Image resolution enhancement using a Hopfield neural network*," Proceedings of IEEE International Conference on Information Technology: New Generations, 2007.

[295] Y. Zhuang, J. Zhang, F.Wu, "*Hallucinating faces: LPH super-resolution and neighbor reconstruction for residue compensation*," Pattern Recognition, vol. 40, no. 11, pp. 3178-3194, 2007.

[296] M.V.W. Zibetti, J. Mayer, "*A robust and computationally efficient simultaneous super-resolution scheme for image sequences*," IEEE Transactions on Circuits and Systems for Video Technology, vol. 17, no. 10, pp. 1288-1300, 2007.

[297] S.W. Park, M. Savvides, "*Robust super-resolution of face images by iterative compensating neighborhood relationships*," Proceedings of the Biometrics Symposium, USA, 2008.

[298] C.V. Jiji, S. Chaudhuri, P. and Chatterjee, "*Single frame image super-resolution: should we process locally or globally?*," Multidimensional Systems Signal Processing, vol. 18, pp. 123-152.

[299] S. Dai, M. Han, W. Xu, Y. Wu, and Y. Gong, "*Soft edge smoothness prior for alpha channel super resolution*," Proceedings of IEEE International Conference on Computer Vision, pp. 1-8, 2007.

[300] M. Ebrahimi and E.R. Vrscay, "*Solving the inverse problem of image zooming using 'self examples*'", International Conference on Image Analysis and Recognition, pp. 117-130, 2007.

[301] Q.X. Yang, R.G. Yang, J.Davis, and D.Nister, "*Spatial-depth super resolution for range images*," Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, USA, 2007.

[302] G.H. Costa, J.C.M. Bermudez, "*Statistical analysis of the LMS algorithm applied to super-resolution image reconstruction*," IEEE Transactions on Signal Processing, vol. 55, no. 5, pp. 2084-2095, 2007.

[303] M. Gevrekci and B.K. Gunturk, "*Super resolution under photometric diversity of images*," EURASIP Journal on Advances in Signal Processing, Article ID 36076, 12 pages, 2007.

[304] G.K. Chantas, N.P. Galatsanos, and N. Woods, "*A super-resolution based on fast registration and maximum a posteriori reconstruction*," IEEE Transactions on Image Processing, vol. 16, no. 7, pp. 1821-1830, 2007.

[305] Y. Yao., B. Abidi, N.D. Kalka, N. Schimid, and M. Adibi, "*Super-resolution for high magnification face images*," Proceedings of the SPIE Defense and Security Symposium, Biometric Technology for Human Identification, 2007.

[306] A.L.D. Martins, M.R.P. Homem, N.D.A. Mascarenhas. "*Super-resolution image reconstruction using the ICM algorithm*," Proceedings of IEEE International Conference on Image Processing, vol. 4, pp. 205-208, USA, 2007.

[307] A. Chakrabarti, A. Rajagopalan, and R. Chellappa, "*Superresolution of face images using kernel pca-based prior*," IEEE Transaction on Multimedia, vol. 9, no. 4, pp. 888-892, 2007.

[308] C.S. Tong and K.T. Leung, "*Super-resolution reconstruction based on linear interpolation of wavelet coefficients*," Multidimensional Systems and Signal Processing, vol. 18, pp. 153-171, 2007.

[309] S. Knorr, M. Kunter and T. Sikora: "*Super-resolution stereo and multi-view synthesis from monocular video sequences*," Proceedings of International Conference on 3-D Digital Imaging and Modeling, pp. 55-64, Canada, 2007.

[310] F.X. Ge, D. Shen, Y. Peng, and V.O.K. Li, "*Super-resolution time delay estimation in multipath environments*," IEEE Transaction on Circuits and Systems, vol. 54, no. 9, pp. 1977-1986, 2007.

[311] F. Lin, C. Fookes, V. Chandran, and S. Sridharan, "*Super-resolved faces for improved face recognition from surveillance video*," ICB, S. W. Lee and S. Z. Li, Eds., Lecture Notes in Computer Science, vol. 4642, pp. 1-10, 2007.

[312] V. Cheung, B.J. Frey, and N. Jojic, "*Video epitomes*," International Journal of Computer Vision, vol. 76, no. 2, pp. 141-152, 2007.
[313] S. Ahmed, N.I. Rao, A. Ghafoor, A.M. Sheri, "*Direct hallucination: direct locality preserving projections (DLPP) for face super-resolution*," Proceedings of IEEE International Conference on Advanced Computer Theory and Engineering, pp. 105-110, Thailand, 2008.
[314] T. Akgun, "*Resolution enhancement using image statistics and multiple aliased observations*," PhD thesis, Georgia Institute of Technology, 2008.
[315] G.M. Callico, S. Lopez, O. Sosa, J.F. Lopez, and R. Sarmiento, "*Analysis of fast block matching motion estimation algorithms for video super-resolution systems*," IEEE Transactions on Consumer Electronics, vol. 54, no. 3, pp. 1430-1438, 2008.
[316] A.S. Beato and G. Pajares, "*Noniterative interpolation-based super-resolution minimizing aliasing in the reconstructed image*," IEEE Transaction on Image Processing, vol. 17, no. 10, pp. 1817-1826, 2008.
[317] F. Brandi, R. L. de Queiroz, and D. Mukherjee, "*Super-resolution of video using key frames and motion estimation*," Proceedings of IEEE International Conference on Image Processing, pp. 321-324, USA, 2008.
[318] G.H. Costa and J.C.M. Bermudez, "*Informed choice of the LMS parameters in super-resolution video reconstruction applications*," IEEE Transactions on Signal Processing, vol. 56, no. 2, pp. 555-564, 2008.
[319] K. Jai, and S. Gong, "*Generalized face super-resolution*," IEEE Transaction on Image Processing, vol. 17, no. 6, pp. 873-886, 2008.
[320] L. Li, Y.D. Wang, "*Face super-resolution using a hybrid model*," Proceedings of IEEE International Conference on Image Processing, pp. 1153-1157, USA, 2008.
[321] Z. Lin, J. He, X. Tang, and C.-K. Tang, "*Limits of learning-based super-resolution algorithms*," International Journal of Computer Vision, vol. 80, no. 3, pp. 406-420, 2008.
[322] R.F. Marcia, and R.M. Willett, "*Compressive coded aperture video reconstruction*," Proceedings of the 16th European Signal Processing Conference, Switzerland, 2008.
[323] R.F. Marcia, C. Kim, J. Kim, D.J. Brady, and R.M. Willett, "*Fast disambiguation of superimposed images for increased field of view*," Proceedings of IEEE International Conference on Image Processing, pp2620-2623, USA, 2008.
[324] J. Mairal, M. Elad, and G. Sapiro, "*Sparse representation for color image restoration*," IEEE Transaction on Image Processing, vol. 17, no. 1, pp. 53-69, 2008.
[325] R. Molina, M. Vega, J. Mateos, and A. Katsaggelos, "*Variational posterior distribution approximation in Bayesian super-resolution reconstruction of multispectral images*," Applied and Computational Harmonic Analysis, vol. 24, no.2, pp. 251-267, 2008.
[326] A. Marquina and S. Osher, "*Image super-resolution by TV-regularization and Bregman iteration*," Journal of Scientific Computing, vol. 37, no. 3, pp. 367-382, 2008.
[327] V.H. Patil, D.S. Bormane, V.S. Pawar, "*Super-resolution using neural network*," Proceedings of IEEE 2nd Asia International Conference on Modelling & Simulation, Malaysia, 2008.
[328] G. Peyre, S. Bougleux, and L. Cohen, "*Non-local regularization of inverse problems*," Proceeding of European Conference on Computer Vision, France, 2008.
[329] P. Sanguansat, "*Face hallucination using bilateral-projection-based two-dimensional principal component analysis*," Proceedings of IEEE International Conference on Computer and Electrical Engineering, pp. 876-880, Thailand, 2008.

[330] Q. Shan, Z. Li, J. Jia, and C.K. Tang, "*Fast image/video upsampling*," Proceedings of ACM Annual Conference Series SIGGRAPH, Computer Graphics, USA, 2008.

[331] K. Simonyan, S. Grishin, D. Vatolin, and D. Popov, "*Fast video super-resolution via classification*," Proceedings of IEEE International Conference on Image Processing, pp349-352, USA, 2008.

[332] F. Sroubek, G. Cristobal, and J. Flusser, "*Simultaneous super-resolution and blind deconvolution*," Journal of Physics: Conference Series, vol. 124, pp. 1-8, 2008.

[333] J. Sun, J. Sun, Z.B. Xx, and H.Y. Shum, "*Image super-resolution using gradient profile prior*," Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, USA, 2008.

[334] X. Wang, J. Liu, J. Qiao, J. Chu and Y. Li, "*Face hallucination based on CSGT and PCA*," Advances in Neural Networks, Lecture Notes in Computer Science, Volume 5264, pp. 410-418, 2008.

[335] Z. Wang, Z. Miao and C. Zhang, "*Extraction of high-resolution face image from low-resolution and variant illumination video Sequences*," Proceedings of International Congress on Image and Signal Processing, China, 2008.

[336] K. Watanabe, Y. Iwai, T. Haga, and M. Yachida: "*A Fast Algorithm of Video Super-Resolution Using Dimensionality Reduction by DCT and Example Selection*," Journal of ITE, vol. 62, no. 11, pp. 1768-1776, 2008.

[337] C.B. Xiao, Y. Jing, X. Yi, "*A high-efficiency super-resolution reconstruction algorithm from image/video sequences*," Proceedings of IEEE 3rd International Conference on Signal-Image Technologies and Internet-based System, pp. 573-580, China, 2008.

[338] Z. Xiong, X. Sun and F. Wu, "*Super-resolution for low quality thumbnail images*," Proceedings of IEEE International Conference on Multimedia & Expo, pp. 181-184, Germany, 2008.

[339] N.A. Yamany, P.E. Papamichalis, "*Robust color image super-resolution: an adaptive M-estimation framework*," EURASIP Journal on Image and Video Processing, Article ID 763254, 12 pages, 2008.

[340] P.H. Yeomans, S. Baker, B. Kumar, "*Recognition of low-resolution faces using multiple still images and multiple cameras*," IEEE International Conference on Biometrics: Theory, Applications and Systems, pp. 1-6, USA, 2008.

[341] P.H. Yeomans, S. Baker, B. Kumar, "*Simultaneous super-resolution and feature extraction for recognition of low-resolution faces*," Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, USA, 2008.

[342] H. Yang, J. Gao, and Z.Wu, "*Blur identification and image super-resolution reconstruction using an approach similar to variable projection*," IEEE Signal Processing Letters, vol.15, pp.289-292, 2008.

[343] J. Yang, J. Wright, T. Huang, and Y. Ma, "*Image super-resolution as sparse representation of raw image patches*," Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, USA, 2008.

[344] M.V.W. Zibetti, F.S.V. Bazan, and J. Mayer, "*Determining the regularization parameters for super-resolution problems*," Signal Processing, vol. 88, pp. 2890-2901, 2008.

[345] L. Baboulaz and P. Dragotti, "*Exact feature extraction using finite rate of innovation principles with an application to image super-resolution*," IEEE Transaction on Image Processing, vol. 18, no. 2, pp. 281-298, 2009.

[346] M. Carcenac, "*A modular neural network for super-resolution of human faces*," Applied Intelligence, vol. 30, no. 2, pp. 168-186, 2009.
[347] T.M. Chan, J.P. Zhang, J. Pu, H. Huang, "*Neighbor embedding based super-resolution algorithm through edge detection and feature selection*," Pattern Recognition Letters, vol. 30, pp. 494-502, 2009.
[348] J.Y. Choi , Y.M. Ro , K.N. Plataniotis, "*Color face recognition for degraded face images*," IEEE Transactions on Systems, Man, and Cybernetics, vol.39, no. 5, pp. 1217-1230, 2009.
[349] G. Costa, and J. Bermudez, "*Registration errors: Are they always bad for super-resolution?*," IEEE Transactions on Signal Processing, vol. 57, no. 10, pp. 3815-3826, 2009.
[350] N. Fan, "*Super-resolution using regularized orthogonal matching Pursuit based on compressed sensing theory in the wavelet domain*," Proceedings of International Conference on Computer Graphics, Imaging and Visualization, pp.349-354, China, 2009.
[351] D. Glasner, S. Bagon, and M. Irani, "*Super-resolution from a single image*," Proceedings of IEEE International Conference on Computer Vision, Japan, 2009.
[352] G. Ginesu, T. Dessì, L. Atzori, and D.D. Giusto, "*Super-resolution reconstruction of video sequences based on back-projection and motion estimation*," Proceedings of International Conference on Mobile Multimedia Communications, UK, 2009.
[353] K. Guo, X. Yang, R. Zhang, S. Yu, "*Learning super resolution with global and local constraints*," Proceedings of IEEE International Conference on Multimedia and Expo, pp. 590-593, USA, 2009.
[354] J.P. Haldar, D. Hernando, L. Zhi-Pei, "*Super-resolution reconstruction of MR image sequences with contrast modeling*," Proceedings of IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pp. 266-269, USA, 2009.
[355] Y. He, K.-H. Yap, L. Chen and L.-P. Chau, "*A soft MAP framework for blind super-resolution image reconstruction*," Image and Vision Computing, vol. 27, pp. 364-373, 2009.
[356] C.C. Hsu, C.W. Lin, C.T. Hsu, H.Y.M. Liao, "*Cooperative face hallucination using multiple references*," Proceedings of IEEE International Conference on Multimedia and Expo, USA, 2009.
[357] H. Ji and C. Fermuller, "*Robust wavelet-based super-resolution reconstruction: theory and algorithm*," IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 31, no. 4, pp. 649-660, 2009.
[358] B. Li and H. Chang, "*Aligning coupled manifolds for face hallucination*," IEEE Signal Processing Letters, vol. 16, no. 11, pp. 957-960, 2009.
[359] X. Li, K.M. Lam, G. Qiu, L. Shen, and S. Wang, "*Example-based image super-resolution with class-specific predictors*," Journal of Visual Communication and Image Representation, vol. 20, no. 5, pp. 312-322, 2009.
[360] B. Li, H. Chang, S. Shan and X. Chen, "*Locality preserving constraints for super-resolution with neighbor embedding*," Proceedings of IEEE International Conference on Image Processing, pp. 1189-1192, Egypt, 2009.
[361] Y. A. Lopez, F. L.-H. Andres, M. R. Pino, and T. K. Sarkar, "*An improved super-resolution source reconstruction method*," IEEE Transaction on Instrumentation and Measurement, vol. 58, no. 11, 3855-3866, 2009.
[362] Y.J. Ma, H. Zhang, Y. Xue, S. Zhang, "*Super-resolution image reconstruction based on K-means-Markov network*," Proceedings of IEEE International Conference on Natural Computation, vol. 1, pp.316-318, China, 2009.

[363] X. Ma, J. Zhang and C. Qi, "*Hallucinating faces: global linear modal based super-resolution and position based residue compensation*," Image Analysis and Processing, Lecture Notes in Computer Science, vol. 5716, pp. 835-843, 2009.

[364] X. Ma, J. Zhang and C. Qi "*An example-based two-step face hallucination method through coefficient learning*," Image Analysis and Recognition, Lecture Notes in Computer Science, vol. 5627, pp. 471-480, 2009.

[365] X. Ma, J. Zhang, C. Qi, "*Position-based face hallucination method*," Proceedings of IEEE International Conference on Multimedia and Expo, pp. 290-293, USA, 2009.

[366] X. Ma, H. Huang, S. Wang, and C. Qi, "*A simple approach to multiview face hallucination*," IEEE Signal Processing Letters, vol. 17, no. 6, pp. 579-582, 2010.

[367] D. Mitzel, T. Pock, T. Schoenemann, and D. Cremers, "*Video super-resolution using duality based TV-L1 optical flow*," DAGM-Symposium, pp. 432-441, 2009.

[368] F. Orieux, T. Rodet, J.-F. Giovannelli, "*Super-resolution with continuous scan shift*," Proceedings of IEEE International Conference on Image Processing, pp. 1169-1172, Egypt, 2009.

[369] V. Patanavijit, "*Super-resolution reconstruction and its future research direction*," AU Journal, vol. 12, no. 3, pp. 149-63, 2009.

[370] D. Patel, S. Chaudhuri, "*Performance analysis for image super-resolution using blur as a cue*," Proceedings of IEEE International Conference on Advances in Pattern Recognition, pp. 73-76, India, 2009.

[371] M. Protter, M. Elad, "*Super-resolution with probabilistic motion estimation*" IEEE Transactions on Image Processing, vol. 18, no. 8, 1899-1904, 2009.

[372] M. Protter, M. Elad, H. Takeda, P. Milanfar, "*Generalizing the nonlocal-means to super-resolution reconstruction*," IEEE Transactions on Image Processing, vol. 18, no. 1, pp. 36-51, 2009.

[373] J. Tian and K-K. Ma "*A state-space super-resolution approach for video reconstruction*," Signal, Image and Video Processing, vol. 3, no. 3, pp. 217-240, 2009.

[374] B. Sankur and H. Ozdemir "*Subjective evaluation of single frame super-resolution algorithms*," Proceedings of European Signal Processing Conference, Scotland, 2009.

[375] P. Sen, S. Darabi, "*Compressive image super-resolution*," Proceedings of 43rd IEEE Asilomar Conference on Signals, Systems and Computers, pp. 1235-1242, USA, 2009.

[376] M. Shao, Y. Wang, Y. Wang, "*A super-resolution based method to synthesize visual images from near infrared*," Proceedings of IEEE International Conference on Image Processing, pp. 2453-2456, Egypt, 2009.

[377] H. Shen, S. Li, "*Hallucinating faces by interpolation and principal component analysis*," Proceedings of International Symposium on Computational Intelligence and Design, pp. 295-298, China, 2009.

[378] R.Z. Shilling, T.Q. Robbie, T. Bailloeul, K. Mewes, R.M. Mersereau, M.E. Brummer, "*A super-resolution framework for 3-D high-resolution and high-contrast imaging using 2-D multislice MRI*," IEEE Transactions on Medical Imaging, vol. 28, no. 5, pp. 633-644, 2009.

[379] E. Turgay, G.B Akar, "*Directionally adaptive super-resolution*," Proceedings of IEEE International Conference on Image Processing, pp. 1201-1204, Egypt, 2009.

[380] N.D. Venkata, N.L. Chang "*Display supersampling*," ACM Transactions on Graphics, vol. 28, no. 1, 2009.

[381] Z. Xiong, X. Sun, and F. Wu, "*Web cartoon video hallucination*," Proceedings of IEEE International Conference on Image Processing, pp. 3941-3944, Egypt, 2009.

[382] J. Yang, D. Schonfeld, "*New results on performance analysis of super-resolution image reconstruction*," Proceedings of IEEE International Conference on Image Processing, pp. 1517-1520, Egypt, 2009.

[383] H. Zhao, Y. Lu, Z. Zhai, "*Example-based facial sketch hallucination*," Proceedings of International Conference on Computational Intelligence and Security, pp. 578-582, China, 2009.

[384] H. Zhao, Y. Lu, Z. Zhai, G. Yang, "*Example-based regularization deployed to face hallucination*," Proceedings of International Conference on Computer Engineering and Technology, vol. 1, pp. 485-489, Singapore, 2009.

[385] H. Takeda, P. Milanfar, M. Protter, and M. Elad, "*Super-resolution without explicit subpixel motion estimation*," IEEE Transactions on Image Processing, vol. 18, no. 9, pp. 1958-1975, 2009.

[386] S.P. Belekos, N.P. Galatsanos, A.K. Katsaggelos, "*Maximum a posteriori video super-resolution using a new multichannel image prior*," IEEE Transactions on Image Processing, vol. 19 , no. 6, pp. 1451-1464, 2010.

[387] D.B. Bhushan, V. Sowmya, K.P. Soman, "*Super-resolution blind reconstruction of low resolution images using framelets based fusion*," Proceedings of International Conference on Recent Trends in Information, Telecommunication and Computing, pp.100-104, India, 2010.

[388] P. Didyk, E. Eisemann, T. Ritschel, K. Myszkowski, H. Seidel "*Apparent display resolution enhancement for moving images*," ACM Transactions on Graphics, vol. 29, no. 4, 2010.

[389] P.P. Gajjar, M.V. Joshi, "*New learning based super-resolution: use of DWT and IGMRF prior*," IEEE Transaction on Image Processing, vol. 19, no. 5, pp. 1201-1213, 2010.

[390] N. Joshi, W. Matusik, E.H. Adelson, D.J. Kriegman, "*Personal photo enhancement using example images*," ACM Transactions on Graphics, vol. 29, no. 12, 2010.

[391] H. Huanga, H. Hea, X. Fanb and J. Zhang, "*Super-resolution of human face image using canonical correlation analysis*," Pattern Recognition, vol. 43, no. 7, pp. 2532-2543, 2010.

[392] P. Gajjar and M. Joshi, "*Zoom based super-resolution: a fast approach using particle swarm optimization*," Image and Signal Processing, Lecture Notes in Computer Science, vol. 6134, pp. 63-70, 2010.

[393] MM. Islam, V.K. Asari, M.N. Islam, and M.A. Karim, "*Super-resolution enhancement technique for low resolution video*," IEEE Transactions on Consumer Electronics, vol. 56, no. 2, pp. 919-924, 2010.

[394] K.I. Kim, Y. Kwon, "*Single-image super-resolution using sparse regression and natural image prior*," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 6, pp. 1127-1133, 2010.

[395] B.G.V. Kumar, R. Aravind, "*Computationally efficient algorithm for face super-resolution using (2D)2-PCA based prior*," IET Image Processing, vol. 4, no. 2, pp. 61-69, 2010.

[396] B. Li, H. Chang, S. Shan, and X. Chen "*Low-resolution face recognition via coupled locality preserving mappings*," IEEE Signal Processing Letters, vol. 17, no. 1, pp. 20-23, 2010.

[397] Y.R. Li, D.Q. Dai, L. Shen, "*Multiframe super-resolution reconstruction using sparse directional regularization*," IEEE Transactions on Circuits and Systems for Video Technology, vol. 20, no. 7, pp. 945-956, 2010.

[398] X. Li, Y. Hu, X. Gao, D. Tao, and B. Ning "*A multi-frame image super-resolution method*," Signal Processing, vol. 90, no. 2, pp. 405-414, 2010.

[399] Z. Liangpei, Z. Hongyan, S. Huanfeng, P. Li, "*A super-resolution reconstruction algorithm for surveillance images*," vol. 90, no. 3, pp.848-859, 2010.

[400] X. Maa, J. Zhang, and C. Qi, "*Hallucinating face by position-patch*," Pattern Recognition, vol. 43, pp. 2224-2236, 2010.

[401] K. Nasrollahi and T.B. Moeslund, "*Finding and improving the key-frames of long video sequences for face recognition*," Proceedings of IEEE Transaction on Biometrics: Theory, Applications and System, USA, 2010.

[402] K. Nasrollahi and T.B. Moeslund, "*Hallucination of super-resolved face images*," Proceedings of IEEE 10th International Conference on Signal Processing, China, 2010.

[403] K. Nasrollahi and T.B. Moeslund, "*Hybrid super-resolution using refined face-logs*," Proceedings of IEEE 2nd International Conference on Image Processing Theory, Tools and Applications, France, 2010.

[404] Q. Pan, C. Gao, N. Liu, and J. Zhu, "*Single frame image super-resolution based on sparse geometric similarity*," Journal of Information and Computational Science, vol. 7, no. 3, pp. 799- 805, 2010.

[405] M.D. Robinson, C.A. Toth, J.Y. Lo, S. Farsiu, "*Efficient Fourier-wavelet Super-resolution*," To appear in IEEE Transaction on Image Processing, 2010.

[406] F. Rousseau, "*A non-local approach for image super-resolution using intermodality priors*," Medical Image Analysis, vol. 14, pp. 594-605, 2010.

[407] M. Shena, C. Wanga, P. Xuea, and W. Lina, "*Performance of reconstruction-based super-resolution with regularization*," To appear in Journal of Visual Communication and Image Representation, 2010.

[408] J. Sun, J.J. Zhu, M.F. Tappen, "*Context-constrained hallucination for image super-resolution*," Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, USA, 2010.

[409] J. Tian, and K.K. Ma, "*Stochastic super-resolution image reconstruction*," Journal of Visual Communication and Image Representation, vol. 21, no. 3, pp. 232-244, 2010.

[410] J. Wang, S. Zhua, and Y. Gonga, "*Resolution enhancement based on learning the sparse association of image patches*," Pattern Recognition Letters vol. 31, pp. 1-10, 2010

[411] Z. Xiong, X. Sun, and F. Wu, "*Robust web image/video super-resolution*," IEEE Transactions on Image Processing, vol. 19, no. 9, pp. 2017-2028, 2010.

[412] J. Yang, J. Wright, T. Huang, and Y. Ma, "*Image super-resolution via sparse representation*," To appear in IEEE Transaction on Image Processing, 2010.

CHAPTER 7

OVER-COMPLETE FACE-LOGS FOR SUPER-RESOLUTION

7 Chapter 7: Over-Complete Face-Logs for Super-Resolution

7.1 Introduction

Faces in surveillance videos are usually very small. This is due to the fact that cameras capturing such videos usually cover a wide field of view and consequently there is a large distance between the objects and the camera. Face recognition systems that work with the facial features that are only extractable in high-resolution frontal face images, have problems in working with low-resolution and low-quality face images coming from such videos. Thus, there is a need for some magnification and enhancement algorithms to improve the quality of the face images before feeding them to the face recognition system.

On the other hand, a one-minute video captured by a surveillance camera working at 30 fps contains 1800 images and applying any enhancement or even face recognition algorithms to all the images of such video is extremely demanding. This is due to the high computational costs of these algorithms. Furthermore, it is neither necessary nor reasonable to apply the recognition or enhancement algorithms to all the images. Usually, there is one or a set of face images that are of better quality compared to their peers and applying the algorithms to them would produce better results in a shorter time than using all the images of the video sequence. These better images are denoted ***key-frames***. Hence, the face quality assessment mechanism described in the previous chapters is needed to summarize an input video sequence to the set of key-frames.

Thus, face recognition systems working with surveillance video sequences should deal with two main problems: reduced reliability of the system (due to working with low-resolution face images) and computational cost of the system (due to the inherent computations of the recognition and enhancement algorithms and also involving all the images of the video sequence). There are many works in the literature to deal with these problems [1-4, to name a few]. Lin et al. [1] use a SR algorithm in conjunction with a face recognition system to enhance the spatial resolution of video frames containing the subject. They have studied the effect of using three different types of SR algorithms on different face recognition algorithms applied to video sequences. Yan et al. [2] used Independent Component Analysis (ICA) to build a liner mixing relationship between high-resolution face image and independent high-resolution source face images. The liner mixing coefficients are retained and the corresponding LR face image is represented by linear mixture of down-sampled source face images. Mixing coefficients to a LR image is obtained by utilizing the unconstrained least square.

For reducing the computational complexity of the enhancement algorithms, some of the authors have tried to do the enhancement processing in another space rather than pixel space. Gunturk et al. [3] have done the SR reconstruction in the eigenface domain. The reconstruction algorithm does not try to obtain a visually improved high-quality image, but instead constructs the information required by the recognition system directly in the low dimensional domain.

Zhifi et al. [4] show that the Two Orthogonal Matrices (TOM) of the Singular Value Decomposition (SVD) of face images contains important information for recognition. Thus, they reconstruct this information instead of performing the SR in the pixel domain.

The mentioned systems achieve acceptable results in reducing the computational time of the system, however, none of these systems nor any other work in the literature have reported their work on ***long video sequences for face recognition***. Either most of these systems use a running window, let say for example of 50 frames, and ***repeat the same process for each window*** or they have reported their results on videos of 2-5 seconds long. The time complexity of using these systems for longer videos of e.g. surveillance cameras is extremely demanding. This is because they treat all the images of the sequence in the same way. Furthermore, the reliability of these systems decrease when they process long video sequences, because they involve non-necessary and/or corrupted data (like non-frontal faces) in the computations.

This chapter is devoted to the last part of the last block of the proposed system shown in Figure 1-1. Face-logs that can be used in different super-resolution algorithms are generated in this chapter using different methods. Meanwhile, a face recognition system is developed to work with long video sequences. To deal with the problem of small size faces, a hybrid super-resolution algorithm and to cope with the computational cost, the face quality assessment system, explained in the previous chapters, are employed, respectively. The system can handle video sequences of any length and resolution. This is because we do not apply nor the enhancement algorithm neither the face recognition to all the frames of the video, instead, we utilize the key-frames of the sequence. We use a window of size 100 frames, find the key-frames, and add them to a face-log. Then, the window is moved to the next segment of the video and the process is repeated. At the end of the video, the enhancement and the recognition algorithms are applied to the face-log and the enhanced result, respectively. Thus, the time-consuming enhancement and recognition algorithms are not applied to each window.

The rest of this chapter is organized as follows: the next section introduces the hybrid super-resolution algorithm that has been developed to work with over-complete face-logs. Section 3 presents a direct and easy, still not so strong method which can construct over-complete face-logs. Then, section 4 resumes the face-log evolution from chapter 5 and generates over-complete face-logs for the new hybrid super-resolution algorithm. Then, sections 5 and 6 explain different versions of this system. Finally, section 7 concludes the chapter.

7.2 Hybrid Super-Resolution

The introduced hybrid super-resolution algorithm uses both types of reconstruction-based and recognition-based super-resolution algorithms. Reconstruction-based super-resolution mainly consists of two important steps: registration of low-resolution images and reconstruction of the high-resolution image [5-13, to new a few]. Registration is of critical importance. Most of the reported super-resolution algorithms are highly sensitive to the errors in registration. Irani and

Peleg [5] are using an iterative method for registration under some assumptions which are only valid for small displacements between the low resolution input images. Due to the movement of objects usually there are large displacements between the first and last appearance of the objects in the real world video sequences. It means that the method of [5] as well as most of the other super-resolution algorithms is not applicable to real video sequences if they are used blindly the same for still images and video sequences.

The following subsections describe the registration algorithm, the reconstruction and the recognition-based super-resolution algorithms, respectively. Two simpler versions of this hybrid super-resolution algorithm are published in [14, 15] and then extended in [16].

7.2.1 Face Image Registration

Having multiple low-resolution input images of the same face, we have used the following registration algorithm from [5]. This registration algorithm takes into account horizontal shift *a*, vertical shift *b* and rotation angle θ between its low-resolution images. Suppose *Y* is the reference image for the registration and X_i is the *i*th low-resolution input image that is going to be registered with *Y*:

This approach takes into account horizontal shift *a*, vertical shift *b* and rotation angle θ between the low-resolution images in the refined face-log. Suppose *Y* is the reference image in the refined face-log and X_i is the *i*th face image in the log and it is going to be registered with *Y*:

$$X_i(k,l) = Y(k\cos(\theta) - l\sin(\theta) + a,\ l\cos(\theta) + k\sin(\theta) + b) \qquad 7\text{-}1$$

Using the first two terms of Taylor's series expansion of $sin(\theta)$ and $cos(\theta)$ gives:

$$X_i(k,l) = Y(k + a - l\theta - \frac{k\theta^2}{2}, l + b + k\theta - \frac{l\theta^2}{2}) \qquad 7\text{-}2$$

Expanding X_i to its own Taylor's series gives:

$$X_i(k,l) = Y(k,l) + (a - l\theta - \frac{k\theta^2}{2})\frac{\partial Y}{\partial k} + (b + k\theta - \frac{l\theta^2}{2})\frac{\partial Y}{\partial l} \quad 7\text{-}3$$

Therefore, the error between X_i and Y can be written as:

$$e(a,b,\theta) = \sum[Y(k,l) + (a - l\theta - \frac{k\theta^2}{2})\frac{\partial Y}{\partial k} + (b + k\theta - \frac{l\theta^2}{2})\frac{\partial Y}{\partial l} - X_i(k,l)]^2 \quad 7\text{-}4$$

The summation is over the overlapping areas of X_i and Y. The minimum of this error can be obtained by computing its derivatives with respect to a, b, and θ and setting them to zero.

$$\begin{aligned} \sum Y_k^2 a + \sum Y_k Y_l b + \sum CY_k\theta &= Y_k D_t \\ \sum Y_k Y_l a + \sum Y_l^2 b + \sum CY_l\theta &= Y_l D_t \\ \sum CY_k a + \sum CYb + \sum C^2\theta &= CD_t \end{aligned} \quad 7\text{-}5$$

where $Y_k=\partial Y/\partial k$, $Y_k=\partial Y/\partial l$, $D_t= X_i-Y$ and $C= kY_l-lY_k$. Solving these equations for a, b, and θ minimizes the difference between the reference image Y and image X_i (ith face image in the refined face-log) warped by (a,b,θ).

Unfortunately, as a critical fact, the above equations can be obtained under assumptions that are valid only for small displacements [5]. It means that using this registration algorithm in any reconstruction-based super-resolution algorithm, which is working with videos, would be extremely erroneous and unreliable. This is due to the differences between the low-resolution inputs that are much more than slight motions. Subsequently, this makes the response of the super-resolution algorithm unstable, noisy and ambiguous. Thus, applying any reconstruction-based super-resolution algorithm directly to the input video sequence, the initial face-log or even to the intermediate face-log (See Figure 5-16 and Figure 5-17) does not produce an acceptable high-resolution output. In section 7.4 it is discussed how we convert the complete face-log of section 5.3.2 into an over-complete face-log which can be used for the following super-resolution algorithm.

7.2.2 Reconstruction-based Super-Resolution

In order to reconstruct a high-resolution image from the low-resolution images of the refined face-log we assume that these images have been produced from a high-resolution image by following an imaging model. Based on the imaging model each low-resolution image has been created by warping, blurring and down-sampling the high-resolution image (Figure 7-1). It means that each X_i, $i=\{1, 2, \ldots, m_3\}$ low-resolution images in the over-complete face-log of section 7.4 have been obtained by the following equation:

$$X_i = DB_iW_iH + n_i \qquad 7\text{-}6$$

where D, B_i and W_i are the down-sampling, blurring and warping matrix, respectively, H is the high-resolution image and n_i is the introduced noise to the imaging process for producing the ith low-resolution image from the high-resolution image H.

Figure 7-1: The imaging model. The desired high-resolution image is at the extreme left and the observed low-resolution image is at the extreme right.

Obtaining the high-resolution image H from equation (7-6) is an inverse problem that is extremely ill-posed and ill-conditioned. Following [6], a MAP method has been employed to obtain the high-resolution response image and a Markov regularization term to convert the problem to a well posed one. Staking all the low-resolution image of the over-complete face-log into a vector $X=[X_1, X_2, \ldots, Xm_3]$, the high-resolution result can be estimated by:

$$\hat{H} = \arg\max\ p(H \mid X) \qquad 7\text{-}7$$

Using the Bayesian rule, it can be written as:

$$\hat{H} = \arg\max \frac{p(X \mid H)p(H)}{p(X)} \tag{7-8}$$

Using the fact that low-resolution images X are known (their probability is constant) and they are independent from each other, Equation 7-8becomes:

$$\hat{H} = \arg\max \prod_{i=1}^{m_3} p(X_i \mid H)p(H) \tag{7-9}$$

Then, using the monotonic logarithm function, we have:

$$\hat{H} = \arg\max \sum_{i=1}^{m_3} [\log p(X_i \mid H) + \log p(H)] \tag{7-10}$$

Now we expand both terms of the above sigma. The image prior $p(H)$ has a Gibbs form:

$$p(H) = \frac{1}{C_3} \exp(-\Gamma(H)) \tag{7-11}$$

where C_3 is a constant and $\Gamma(H)$ is the prior energy function which works as the regularization term. Assuming the noise is i.i.d in Equation 7-6, the likelihood distribution $p(Xi|H)$ in Equation (18) can be expressed as:

$$\log p(X_i \mid H) = \frac{1}{C_4} \exp(-\frac{\| X_i - DB_iW_iH \|^2}{2\sigma_i^2}) \tag{7-12}$$

where C_4 is a constant and σ^2_i is the error variance. The MAP response of the system can be obtained by substituting 7-11 and 7-12 into 7-10 which gives:

$$\hat{H} = \arg\max \sum_{i=1}^{m_3} [\| X_i - DB_iW_iH \|^2 + \lambda\Gamma(H)] \qquad 7\text{-}13$$

where λ is the regularization parameter. For more details regarding the regularization method, the blur and color considerations, the reader is motivated to see [5], [4], and [6], respectively. The results of applying this algorithm to the over-complete face-log is shown in section 7.4. This method improves the quality of the low-resolution images in the refined face-log by a factor close to two. For further improvement of this high-resolution image, it is fed to the next step of the system that is a recognition-based super-resolution algorithm.

7.2.3 Recognition-based Super-Resolution

The employed recognition-based super-resolution algorithm in this paper is a Multi Layer Perceptron (MLP). The benefit of using neural networks is their auto-ability in learning the complicated space of the human faces. For training this MLP we have used around 400 frontal and semi-frontal face images. Some of these images (around 90) are from the same dataset that we use for training the auto-associated networks in the head-pose estimation in section 4.2.1.1 and the others are from CalTech face database (DB1). To prepare the data for training the network, first, we have extracted the face areas of the high-resolution images and converted them to gray-scale. Then, a low-resolution image is created for each of these high-resolution images by down-sampling the images by a factor of two. Then, all of these low-resolution / high-resolution pairs are fed to the network as the training samples and the network learns the relationship between them. The original high-resolution face images in this database are all resized to 184x224 and their corresponding low-resolution images are of 92x112 pixels. This three-layered MLP has 92x112 neurons in its input layer, each corresponding to one of the pixels of the input LR image, 10 neurons in the hidden layer, and 184x224 neurons in the output layer, each corresponding to one of the pixels of the output high-resolution image. The result of applying this algorithm to the result of the reconstruction-based super-resolution of the previous section is shown in section 7.4.

Now that the idea of the hybrid super-resolution has been introduced, in the following sections we discuss different systems that have been developed for the purpose of the constructing over-complete face-logs. Except the system explained in the next two sections, the other systems apply a super-resolution algorithm similar to the above explained one to their over-complete face-logs.

7.3 OCFL-System 1

Our sixth proposed system (OCFL-System 1) [16] uses the four facial features explained in 4.2.1.2, 4.2.1.3, 4.2.1.4, and 4.2.1.5, respectively. For combining the normalized value of the quality measures and producing a quality score for each face image, OCFL-System 1 uses a fuzzy inference engine similar to the one explained in 5.3.2.1. Having extracted the quality scores for the face images a quality curve is created for the video sequence. Such a quality curve is shown in Figure 7-2 for a given video sequence.

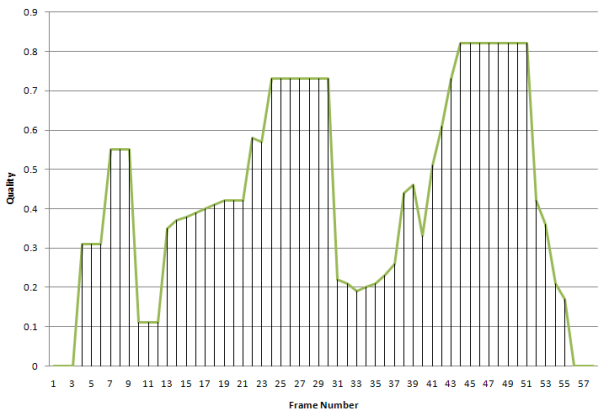


Figure 7-2: Quality curve for a given sequence (quality vs. frame number).

Instead of applying the super-resolution algorithm to the entire video sequence, the peaks of this quality curve are found and a face-log is generated for each of these peaks, then the super-resolution algorithm is applied to these logs, separately. The images inside each face-log are temporally close to each other and as the quality curve shows they are similar in terms of quality. Thus, their registration error would be much smaller than the registration of the entire video sequence.

Usually building *m* (*m*=3-5) face-logs corresponding to the *m* highest peaks suffices. However, theoretically the number of the face-logs can be equal to the number of the local maxima in the quality curve. Each face-log contains all the images located on a specific local maximum. For the quality curve shown in Figure 7-2, the face-logs corresponding to the three highest peaks of this curve are shown in Figure 7-3(a)-(c). Each local maximum contains images which are more or less similar in terms of quality. These images are temporally close to each other, i.e. rather than similarity in quality except some small variations in motion, lightening, rotation and scales they resemble each other closely and so the compensation between them can be done more easily and accurately which makes the super-resolution algorithm more robust and fast.

Figure 7-3: Face-logs corresponding to the three highest peaks of the quality curve in Figure 7-2.

By this way of face-log constructing one can be sure that useless images which are the source for most of the errors in registration step of super-resolution are ignored and faces with more spatial similarity are considered together. This spatial similarity dose not prevents images of having slight differences in rotation, scale, brightness, and motion which are necessary for super-resolution.

This system uses the super-resolution algorithm developed by Zomet and Peleg [11]. This super-resolution algorithm follows the image model that is formulated in [12]. For experimental results 50 video sequences from DB7 are used. Two tests have been performed for each of these sequences: first applying the super-resolution algorithm to the face-logs generated by the above mentioned method and second applying the super-resolution-algorithm directly to the video sequences.

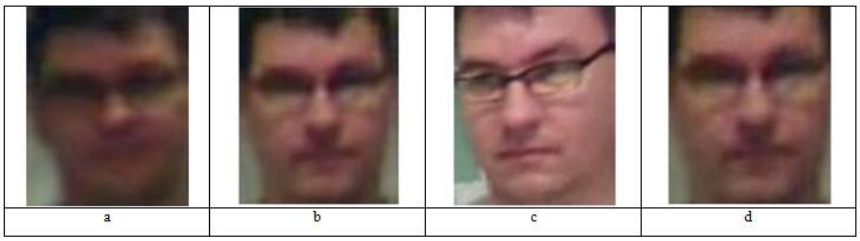


Figure 7-4: Results of applying the super-resolution to a) first, b) second, c) third face-log of the sequence of Figure 7-3. d) Result of the algorithm applied to all the faces in that sequence.

Figure 7-4 shows the results of the both tests for the sequence given in Figure 7-3. As can be seen from this figure applying the super-resolution to the face-logs generates much better results than applying it directly to the video sequences. It means that the qualities of the images generated by the first method are much better than their counterparts generated by the second method. First images reveal more details of the face and obviously are more useful than the

ones generated by the second method. The reason is that compensating for the changes inside a face-log is much easier with fewer errors than doing that for the whole sequence.

7.4 OCFL-System 2

This section continues the unfinished discussion of section 5.3. Having constructed the best face-log and the complete face-log from a low-resolution input video sequence, now it is the time to evolve these face-logs for using in super-resolution algorithms.

Reconstruction-based super-resolution algorithms, and consequently our hybrid super-resolution algorithm, can produce the missing high-resolution details correctly if they have enough low-resolution input images from the same face. Furthermore, the low-resolution input images of these algorithms should only have sub-pixel misalignments between themselves. These misalignments, however, should be considered in a registration step before getting into the actual reconstruction of the high-resolution output image.

Our seventh proposed system (OCFL-System 2) [17] uses the superset of facial features and combines them using a weighting system to produce a quality score for each face image to BFI-System 1 in section 5.3.1.1. This system uses the previously generated intermediate face-logs (Intermediate face-logs are generated from the initial face-logs by choosing the *m*-best face-image of those logs, Figure 5-16 and Figure 5-17.). We pick the best face images (frontal and side-views) of the intermediate face-logs as the reference frame for that log. Then, for each intermediate face-log, two similarity measures are calculated between the members of the logs and the reference frame of the log.

The two used similarity measures are Correlation Coefficient and Structural Similarity Measure [19]. For calculating these similarity measures, we need to resize the images to the same size as the reference image and convert them to gray-scale. Correlation coefficient score (7-14) shows the relationship between an image in the intermediate face-log and the best image (reference image) of the same log, in terms of the least squares.

$$SM_{1X_i} = \frac{\sum_{k=1}^{K}\sum_{l=1}^{L}[(X_i(k,l),-avg(X_i))(Y(k,l),-avg(Y))]}{\sqrt{\sum_{k=1}^{K}\sum_{l=1}^{L}[X_i(k,l),-avg(X_i)]}\sqrt{\sum_{k=1}^{K}\sum_{l=1}^{L}[Y(k,l),-avg(Y)]}} \qquad 7\text{-}14$$

where X_i is the *i*th face image in the intermediate face-log, Y is the reference face image in this log (the best face image in the log), SM_{1X_i} is the first similarity measure between these two

images, *avg* is a function that returns the average value of its parameter, and *k* and *l* change across the dimensions of the images.

Structural similarity (7-15) measure takes into account contrast, luminance and the structure of the images to determine their similarity [19]. It is defined as:

$$SM_{2X_i} = avg[\frac{(2avg(x)avg(y)+C_1)(2\tau(x,y)+C_2)}{(avg(x)^2+avg(y)^2+C_1)(\sigma(x)^2+\sigma(y)^2+C_2)}] \quad 7\text{-}15$$

where C_1 and C_2 are constant, $SM_{2X_i}$is the second similarity measure between the *i*th face image in the intermediate face-log and the reference face image, *x* and *y* are sub-images of these two images, σ returns the standard deviation of its parameter and τ is defined as:

$$\tau(x,y) = \frac{1}{KL-1}\sum_{k=1}^{K}\sum_{l=1}^{L}[(x(k,l),-avg(x))(y(k,l),-avg(y))] \quad 7\text{-}16$$

The maximum values for both correlation coefficient and structural similarity measures are one, which indicate the perfect correlation and similarity, respectively, between the *i*th face image in the intermediate face-log and the reference face image of that log.

$$SM_{X_i} = \frac{SM_{1X_i}+SM_{2X_i}}{2} \quad 7\text{-}17$$

The mean of these two similarity measures (7-17) is then considered as the similarity factor between the current face and the reference face. Based on this similarity factor the n_4-most ($n_4<n_3$) similar images to the reference image for each log are chosen and considered as the member of what is denoted an over-complete face-log. Figure 5-16(d) and Figure 5-17(d) show the three over-complete face-logs generated for the input video sequence given in Figure 5-16(a) and Figure 5-17(a), respectively. This over-complete face-logs are then used as the input to the super-resolution algorithm.

7.4.1 OCFL-System 2: Experimental Results

This system is tested using four databases: FRI CVL (DB3), AT&T (DB5), Face96 (DB6), and the Local database (DB10). Several tests have been done on the system. In the first test, the proposed system tries to find the best face image(s) of the input video or image sequence. Having obtained the quality score for each image of the sequence, a ranking number is assigned to each face image. This ranking number is increased as the quality score decreases. To check the performance of the system in finding the best image(s) of the given sequence, we have asked an expert to annotate the images based on his perception of the quality of the images and assign a ranking number to each figure. Then, the results of the system are compared against the annotated data (ground truth). To validate the annotation done by the first expert, we have asked more than 10 other experts to do the same process on random sequences (totally 85 sequences) from all the databases. In more than 97% of the cases the results of the second group of experts is in complete agreement with the results of the first expert. Thus, we consider the annotated data prepared by the first expert as the ground truth. Similar, to the previous systems, the results of this system are generally in good agreement with the ground truth.

In the second test, the system constructs a complete face-log. By complete, we mean that such a face-log is enough for identification purposes. This log should contain the most frontal and two different side-view face images. Thus, the minimum number of the face images in this log is three. The experiment carried out for the first test cannot construct such a face-log with the minimum number of members. If we choose the three images with highest quality score for the face-log, the log would contain the best image and two other image that are similar to it not the side-view images. Thus, as explained 5.3.2.2, we first use the head's pan information to divide the input video sequence into three logs and then find the best images of each log to construct a complete face-log. Figure 5-16(c) and Figure 5-17(c) show the members of a complete face-log generated for the video sequence given in Figure 5-16(a) and Figure 5-17(a), respectively.

If the original input video sequence is of high quality, such a complete face-log is perfect for identification purposes. However, if the input video sequence is of low quality and low-resolution, as it is the case in the surveillance scenarios, even such a complete face-log may not be useful without some enhancement algorithm. Therefore, we use the reconstruction-based super-resolution algorithm of [6]. In the third test, the system evolves the complete face-log(s) of the second test to prepare suitable inputs for this super-resolution algorithm. Figure 5-16(d) and Figure 5-17(d) show the over-complete face-logs generated for this purposes for the video sequence given in Figure 5-16(a) and Figure 5-17(a), respectively.

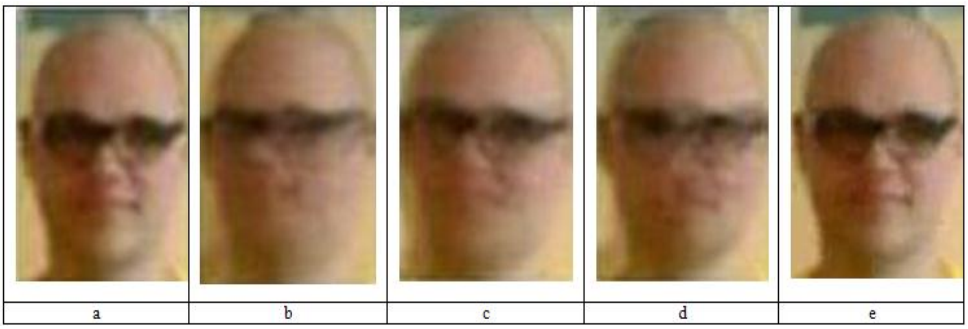


Figure 7-5: the best frontal face image of the input video sequence given in Figure 5-16(a), and results of the super-resolution algorithm applied to: b) all the images of the input video sequence, c) the initial frontal face-log given in Figure 5-16(c2), d) the intermediate face-log of the sequence (which is the same as the log shown in Figure 5-16(b) for this sequence), and finally e) the over-complete frontal face-log of the sequence that are the images shown in Figure 5-16(d2).

Figure 7-5 shows the results of applying the employed super-resolution algorithm to the frontal face images of Figure 5-16(d2). It can be seen that the facial details are more visible in the enhanced images than in the complete log. In order to show the effectiveness of the used face quality assessment technique on the employed enhancement algorithm, the algorithm has been applied to the intermediate results of the assessment technique, i.e. the different face-logs. Figure 7-5(b)-Figure 7-5(d) show the result of the enhancement algorithm applied to the entire video sequence of Figure 5-16(a), the initial, and the intimidate frontal face-logs, respectively. These images should be compared to the image of Figure 7-5(e). It can be seen that applying the super-resolution algorithm to the different logs of the sequence produces results that are worse than the best image of the sequence (Figure 7-5(b)-Figure 7-5(d)). However, the result of using the super-resolution algorithm for the refined (over-complete) face-log is much better than the other results and slightly better than the best image of the sequence. It is more visible in the other example shown in Figure 5-17 and its results shown in Figure 7-6.

The explained process for summarizing input video sequences to different types of face-logs has been repeated for more than 100 real-video sequences. In most of the cases, the proposed system can find the best both frontal and side-view face images, very well. Thus, the result of this part of the system can be used for video indexing, reliably. Since the system can find the best frontal and side-view face images, the complete face-log, containing these three face images can be constructed. Such complete face-logs are enough for identification purposes. The used similarity measures can evolve the complete face-logs to logs that are suitable for super-resolution algorithms.

Figure 7-6: a) the best frontal face image of the input video sequence given in Figure 5-17(a), and results of the super-resolution algorithm applied to: b) all the images of the input video sequence, c) the initial frontal face-log given in Figure 5-17(c2), d) the intermediate face-log of the sequence (which is the same as the log shown in Figure 5-17(b) for this sequence), and finally e) the refined (over-complete) frontal face-log of the sequence that are the images shown in Figure 5-17(d2).

7.5 OCFL-System 3

Our eighth proposed system (OCFL-System 3) [16], as the previous system uses the combination of quality measures, pan information, and similarity measures to construct the best face-log, complete, and over-complete face-logs, respectively (See Figure 7-7). It employs the hybrid super-resolution algorithm introduced in the beginning of this chapter. The four first facial features of the superset of facial features are used and they are combined using a weighting system. The weights obtained for these features are shown in Table 7-1.

Table 7-1: Weights of the facial features involved in the quality assessment of OCFL-System3

Features	*Value of the weights*
Head-pose	1.7
Sharpness	0.9
Brightness	0.6
Resolution	0.8

To obtain the weights of the features shown in Table 7-1 we have annotated the best images of 30 video sequences of DB10, manually. Then, the system has been applied to these videos using different values of the weights of the features as shown in Table 7-2. For each set of values, we have repeated the assessment process and monitored the matching rate between the results of the algorithm and the annotated data. The matching rates for different values of the weights are shown in Table 7-2. It can be seen from this table that the best matching rate is obtained when the values of the features are the same as the values shown in Table 7-1.

Table 7-2: Changing the weights of the four features involved in the quality assessment to obtain a proper set of weights and the matching rate between the result of the quality assessment and the annotated data for finding the best images of 30 video sequences from DB10.

	Head-pose						
Sharpness		0.5	0.9	1.3	1.7	2.1	Resolution
	0.3	0.41	0.69	0.61	0.77	0.82	0.4
	0.6	0.53	0.65	0.65	0.78	0.81	0.6
	0.9	0.59	0.68	0.76	0.85	0.81	0.8
	1.2	0.61	0.72	0.76	0.82	0.83	1
	1.5	0.59	0.71	0.79	0.82	0.83	1.2
		0.3	0.4	0.5	0.6	0.7	
	Brightness						

7.5.1 OCFL-System 3: Experimental Results

To show the efficiency of the proposed system in real-world situations, it has been tested using two different databases. The first one is the local database (DB10) and the second one (DB11), NRC-IIT [20], is a publicly available database containing 22 low-resolution video sequences of 11 test-subjects. People in this database are sitting on a movable chair in front of a camera and change their head-pose, facial expressions, and distance to camera. The average length of the videos of this database is around 14 seconds.

Five different experiments have been carried out on these video sequences. The first experiment validates the performance of the cascaded super-resolution algorithms over the separate using of the employed super-resolution algorithms. Figure 7-7(l) shows the results of applying the proposed system to the refined (over-complete) face-log of the video sequence in Figure 7-7(a). Figure 7-7(j) shows the result of applying the reconstruction-based super-

a

B c

d

Continues to the next page

Figure 7-7: a) Every *m*th frame (*3<m<15*) of a video sequence from DB10 and two different face-logs of this video which are produced for different purposes: b) for video indexing and c) for summarizing the video sequence (complete face-log). Based on the value of the head-pose: d) initial frontal face-log, e) initial left side-view face-log, and f) initial right side-view face-log, g) the intermediate, and h) the refined (over-complete) frontal face-log for the frontal face-log, i) The best face image of the video sequence j) result of the reconstruction-based super-resolution for the refined face-log of that sequence, k) result of the recognition-based super-resolution for the best image of the sequence, l) result of the proposed system, m) result of reusing the reconstruction-based algorithm applied to j), n) result of applying the system to the intermediate face-log, and o) result of applying the system to the initial face-log.

resolution algorithm on the refined face-log. Figure 7 7(k) shows the result of applying the recognition-based super-resolution algorithm to the best image of the same refined face-log. Figure 7 7(m) shows the result of reusing the reconstruction-based super-resolution algorithm for further improvement of the previous result. From these figures (Figure 7 7(a)-Figure 7 7(m)) it can be seen that cascading our super-resolution algorithms (Figure 7 7(l)) can produce better results than using them separately, or using them for the second time.

The second experiment illustrates the importance of the used face quality assessment technique. Figure 7 7(n) and Figure 7 7(o) show the results of applying the proposed system to the intermediate face-log and the initial face-log, respectively. These images should be compared with the result of applying the system to the refined face-log, shown in Figure 7 7(l). It can be seen that using the face quality assessment and the face-log generation techniques are the reason for the better result of the latter case. This is due to classifying similar images in the same class and thus reducing the registration error. This consequently improves the final response of the super-resolution system.

The third experiment shows the importance of choosing the reference frame in the reconstruction-based super-resolution part. Second row of Figure 7-8 shows the results of the

proposed system using the images shown on the first row of the same figure as the reference frame of the system. Even though in all of these cases the face-log is the same, the result of the system changes according to the changes in the reference frame. The system tries to produce the high-resolution output such that it resembles the reference frame. Therefore, it is very important to choose the best image of the intermediate face-log as the reference frame in the refined face-log. Moreover, since we want to produce a ***frontal*** high-resolution face image as the output of the system, it is critical that the best-chosen image (or reference image) is the least rotated image among its peers in the initial face-log. This is the reason for the bigger weight of the head-pose compared to the other features in the quality assessment part.

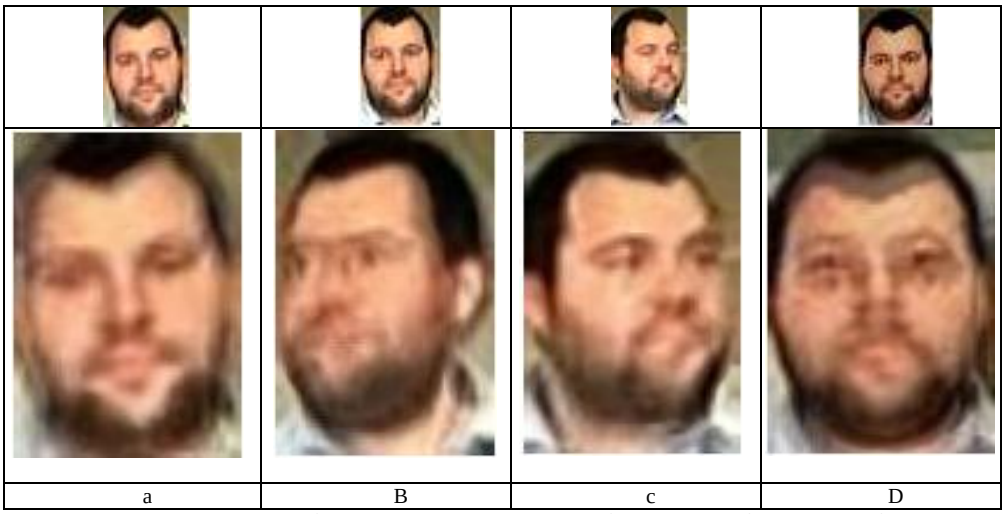


Figure 7-8: The importance of choosing the best image as the reference image: If images in the first row are chosen as the reference image, the output of the system would be as the second row.

The fourth experiment shows the real impact of the proposed system in improving the recognition rate of a linear associative face recognizer. The recognizer is trained using the (manually extracted) best face images of the video sequences of both databases. To do the experiment, the best images of the sequences are first found by the face quality assessment algorithm. Then, they are resized to the required size of the inputs of the recognition algorithm by both bilinear and bicubic algorithms. Then, the recognition rate of the face recognizer is monitored in these cases and compared against the case that the input of the recognizer is produced by the proposed system. The recognition rates are shown in Table 7-3. It can be seen that the face recognizer performs better when it is fed using the high-resolution images produced by the proposed system.

Table 7-3: Improving the recognition rate of a linear auto-associative face recognizer when its input is prepared by the proposed method rather than by bilinear or bicubic algorithms.

Method	*Recognition Rate*
Bilinear	53.6
Bicubic	57.1
The proposed system	86.7

The fifth experiment shows the generalization capability of the system for non-frontal face images. As discussed before, the actual numbers of the refined face-logs are three, one for the frontal face images, and two for the side-views (left and right) face images. The refined side-view face-logs can be used for generating high-resolution side-view face images of the objects. The side-view high-resolution output images obtained for the video sequence in Figure 7-7(a) are shown in Figure 7-9(a) and Figure 7-9(b). An application using three such high-resolution images (one frontal and two side-view images) is 3D head model generation. We have applied these three high-resolution images to the software available at [21] to create a 3D model for the person. Figure 7-9(c) shows this model.

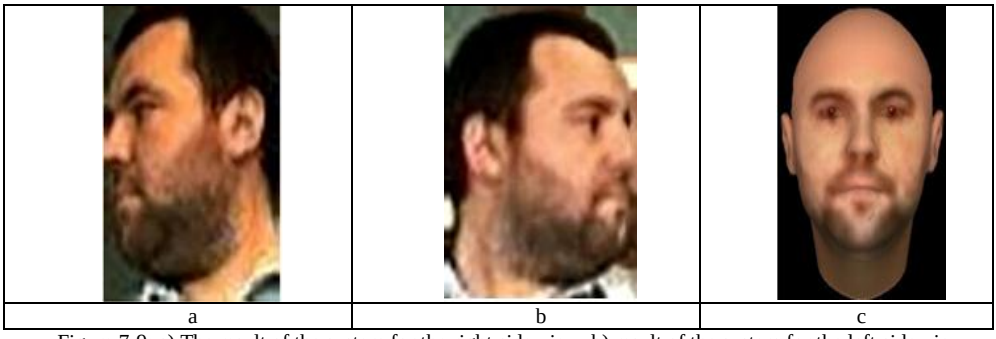


Figure 7-9: a) The result of the system for the right side-view, b) result of the system for the left side-view, and c) the 3d model of the face.

Figure 7-10 shows the required processes from detecting the faces to producing the result of the proposed system for another video sequence from DB10. Figure 7-10(a) shows the faces detected from the input video sequence. Figure 7-10(b) shows the initial frontal face-log obtained from Figure 7-10(a). Figure 7-10(c) and Figure 7-10(d) show the intermediate and refined frontal face-log, respectively. Figure 7-10(e) shows the results of the above-explained experimental tests for this video sequence.

a) every *m*th *(3<m<15)* frame of an input video sequence.

b) the initial frontal face-log.

c) the intermediate frontal face-log.

d) the refined frontal face-log.

e) from left to right: best face image of the input video sequence, result of the reconstruction-based SR, result of the recognition-based SR, result of the proposed system (combintion of both SR algorithms), result of the reconstruction-based algorithm applied to the second image from left, result of the system for the intermediate frontal face-log and finally, the result of the system for the initial frontal face-log.

Figure 7-10: Obtaining the HR frontal image for another video sequence from DB10.

Figure 7-11 and Figure 7-12 show the results of the system for two more video sequences from DB10. The system fails to produce the high-resolution frontal image in Figure 7-12. This is due to the wide changes in lighting conditions while capturing its associated video sequence. It makes the facial feature extraction erroneous and consequently causes the face quality assessment to fail in finding the best face image of the sequence and construction the face-logs, correctly. Therefore, the result of the system is noisy and unstable. Figure 7-13 shows the same images for another video sequence from DB11.

Figure 7-11: The results of the system for another video sequence from DB1, for descriptions see Figure 7-10(e).

Figure 7-12: The results of the system for another video sequence from DB10 where the system fails to produce a frontal HR image. For descriptions of the images, see Figure 7-10(e).

Figure 7-13: The results of the system for a video sequence from DB11. See Figure 7-10(e) for descriptions of the images.

7.6 OCFL-System 4

Our ninth proposed system (OCFL-System 4) [22] is also similar to OCFL-System 2 in terms of number and types of the involved facial features, their combination method, and the face-logs construction algorithms. This system is mainly developed for finding and improving the key-frames of long video sequences for face recognitions. These key-frames are found by the face quality assessment part and then the similarity measures are used to find their similar peers to be fed all together to the hybrid super-resolution algorithm. The key point of this system is that faces of small sizes have been used for carrying out the experimental results. The images in FRI CVL (DB3), Face96 (DB6), and the Local database (DB10), and 50 sequences from the public available database FERET (DB12) [23] are down-sampled by a factor of 3 and then they are fed to the system. The total number of sequences from these databases are 450 sequences (FERET (50), FRI CVL (144), Face96 (152), and local database (120)). This number of small sizes sequences ensures that the proposed method for over-complete face-log construction can be efficiently applied to face recognition system when their inputs are of small sizes.

7.6.1 OCFL-System 4: Experimental Results

Several experimental results have been performed using these low-resolution face images. The first experiment shows that the system can find the key-frames, which are the best images of the sequences, even in these low-resolution sequences. To do so, we again compare the results of the system against the annotated data. This is shown in Figure 7-14 for three sequences from the used public databases.

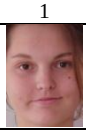
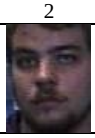
				
4	2	1	3	4
4	2	1	3	5
				
5	4	1	2	3
4	3	1	1	2
				
5	3	2	4	1
5	3	4	2	1

Figure 7-14: Some images of three different sequences from three public databases: FERET, FRI CVL (mid.), and Face96. The ground truth (first row of each sequence) vs. ranking numbers given by the system (second row).

Even though the results of the system in general are not as good as BFI-System 1 or BFI-System 2, they are still in good agreement with the annotated data. Figure 7-15 shows the overall results of agreements between the ranking numbers given by the system and the ground truth for finding the first, second, and third best images of all the sequences of different databases.

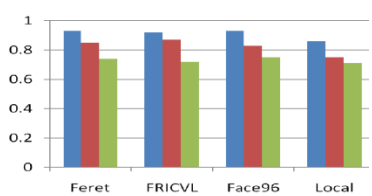


Figure 7-15: The overall agreement between the ground truth and the proposed system in finding the first, the second, and the third best images, respectively, from four different databases.

The second experiment shows that the enhanced image produced by the system is better than the best image of the video sequence. Furthermore, applying the hybrid super-resolution of this chapter to the over-complete face-log of the best key-frames produces a much better results than applying it to the entire sequence or even to the intermediate face-logs. Figure 7-16(a) shows an input video sequence and summarizing it to a face-log containing the most frontal face images in part (b), then the best key-frames in part (c) and finally to the reference key-frame and the over-complete face-log in part (d) of the figure.

a) Every *x*th (*4<x<16*) frame an input video sequence

b) The face-log containing the *m* most frontal face images of the sequence

c) The face-log containing the *n* (*n<m*) best key-frame of the sequence

d) The reference key-frame (left), the refined (over-complete) face-log containing the best key-frames of the sequence (right).

Figure 7-16: Obtaining the key-frames of a given video sequence.

The result of the super-resolution applied to the entire sequence of Figure 7-16(a) and its intermediate face-log (Figure 7-16(b)) are shown in Figure 7-17(d) and Figure 7-17(e), respectively. This image should be compared to the image obtained from applying the super-resolution algorithm to the refined (over-complete) face-log of the key-frames shown in Figure 7-17(c).

A	b	c	d	e

Figure 7-17: Improving the quality of the best image of the sequence given in super-resolution Figure 7-16: a) the best image of the sequence, b) the result of the reconstruction-based super-resolution applied to the face-log of the key-frames shown in Figure 7-16(d), c) the result of applying the recognition-based super-resolution applied to the previous image (The grayscale version of this image is fed to the recognition algorithm) and finally, d and e show that applying the super-resolution algorithm to the refined face-log of the key-frames is much better than applying it to the entire sequence (d) shown in Figure 7-16(a) or even the intermediate face-log (e) shown in Figure 7-16(c).

The proposed system employs a face recognition algorithm that needs its input to be a gray-scale image of size 92x112 pixels. In the third and last experiment, we compare our system with some well-known approaches in the literature to see which one can prepare a better input to the face recognition algorithm and subsequently can improve the recognition rate of the system. To do so, we have obtained the input of the recognition algorithm from the key-frame of the sequence (and/or the refined face-log containing the key-frame and its similar peers) using the methods listed in the first column of Table 7-4. Then, the result of the used recognition algorithm is monitored. This has been repeated for all the video sequences of the locally prepared database (DB10). This database contains 120 video sequences captured from 30 different persons. People in these videos have been asked to walk around in front of a Logitech camera, freely. It is reported in the second row of Table 7-4 that the proposed system can yield the best recognition rate among the studied methods.

Table 7-4: Comparison against the similar systemsin the litrature: changes in the recognition rate of the employed face recognition system when its input is prepared by: a) bilinear, b) cubic spline interpolation, c) Schultz et. al.[24], d) Baker et. al. [25], and e) the proposed system.

Method	*a*	*b*	*c*	*d*	*e*
Recognition Rate	51.6	53.2	71.4	79.2	85.3

7.7 Summary

This chapter continues the discussion of face-log generation from chapter 5 and converts the complete face-logs into over-complete ones that are suitable for our hybrid super-resolution algorithm. This super-resolution algorithm is also introduced in this chapter and its efficiency when it is applied to the over-complete face-logs are shown. Different methods for generation

such face-logs are discussed in this chapter. Finally, a face recognition algorithm is developed and it is shown that the inputs obtained from the over-complete face-logs using the hybrid super-resolution algorithm can improve the recognition rate of this system.

References

[413] F. Lin, C. Fookes, V. Chandran, and S. Sridharan, "*Super-Resolved Faces for Improved Face Recognition from Surveillance Video*," The 2nd International Conference on Biometrics, Korea, LNCS, 2007, pp. 1-10.

[414] H. Yan, J. Liu, J. Sun and X. Sun, "*ICA-based super-resolution face hallucination and recognition*," 4th International Symposium on Neural Networks China, LNCS, 2007, pp.1065-1071.

[415] B.K. Gunturk, A.U. Batur, Y. Altunbasak, M. Hayes, and R.M. Mersereau, "*Eigenface domain Super-Resolution for Face Recognition*," IEEE Trans. Image Processing, vol. 12, no. 5, 2003, pp. 597-606.

[416] W. Zhifei and M. Zhenjiang, "*Feature-based Super-Resolution for Face Recognition*," IEEE International Conference on Multimedia & Expo, Germany, *2008*, pp.1569-1572.

[417] M. Irani, and S. Peleg, "*Improving Resolution by Image Registration*," Graphical Models and Image Processing, vol. 53, no. 3, 1999.

[418] L. Zhang, H. Zhang, H. Shen and P. Li, "*A Super-Resolution Reconstruction Algorithm for Surveillance Images*," International Journal on Signal Processing, vol. 90, no. 3, pp.848-859, 2010.

[419] J. Yang, J. Wright, T. Huang and Y. Ma, "*Image Super-Resolution as Sparse Representation of Raw Image Patches*", International Conference on Computer Vision and Pattern Recognition, USA, 2008.

[420] V. Bannore, "Iterative Interpolation Super-Resolution Image Reconstruction," Springer-Verlag Berlin Heidelberg, 2009.

[421] S. Chaudhuri, "*Super-Resolution Imaging*," Kluwer Academic Publishers, 2nd edition, New York, 2002.

[422] S. Chaudhuri, and M.V. Joshi, "*Motion Free Super-Resolution*," Springer Science, New York, 2005.

[423] A. Zomet, and S. Peleg, "*Super-Resolution from Multiple Images Having Arbitrary Mutual Motion*," In: S. Chaudhuri, Editor, Super-Resolution Imaging, Kluwer Academic, Norwell, pp. 195–209, 2001.

[424] M. Elad, and A. Feuer, "*Super-Resolution Reconstruction of Image Sequences*," IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 21, no. 9. pp. 817-834, 1999.

[425] S. Baker, and T. Kanade, "*Limits on Super-Resolution and How to Break Them*," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 9, 2000.

[426] K. Nasrollahi, and T.B. Moeslund, "Hallucination of Super-resolved Face-Images," IEEE 10th International Conference on Signal Processing, China, 2010.

[427] K. Nasrollahi, and T.B. Moeslund, "Hybrid Super-Resolution using Refined Face-Logs," IEEE 2nd International Conference on Image Processing Theory, Tools and Applications, France, 2010.

[428] K. Nasrollahi, and T.B. Moeslund, "Extracting a Good Quality Frontal Face Image from a Low-Resolution Video Sequence," under review, IEEE Transaction on Circuits and Systems for Video Technology.
[429] K. Nasrollahi, and T.B. Moeslund, "Face-log Generation for Super-Resolution Using Local Maxima in the Quality Curve," International Conference on Computer Vision Theory and Applications, France, 2010.
[430] K. Nasrollahi, T.B. Moeslund, and Mohammad Rahmati "*Summarization of Surveillance Video Sequences using Face Quality Assessment*," To appear, International Journal of Image and Graphics, January 2011.
[431] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "*Image Quality Assessment: From Error Visibility to Structural Similarity*," IEEE Transaction on Image Processing, vol. 13, no. 4, pp. 600-612, 2004.
[432] D.O. Gorodnichy, "*Video-based Framework for Face Recognition in Video*," 2nd Canadian Conference on Computer and Robot Vision, Canada, 2005.
[433] Facegen Modeler, available at: http://www.facegen.com/modeller.htm
[434] K. Nasrollahi, T.B. Moeslund, "*Finding and Improving the Key-frames of Long Video Sequences for Face Recognition*," 4th IEEE International Conference on Biometrics: Theory Applications and Systems, 2010.
[435] P. J. Phillips, H. Moon, P. J. Rauss, and S. Rizvi, "*The FERET Evaluation Methodology for Face Recognition Algorithms*," IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 22, no. 10, 2000, pp. 1090-1104.
[436] R. Schultz and R. Stevenson, "*Extraction of High-Resolution Frames from Video Sequences*," IEEE Transaction on Image Processing, vol. 5, no. 6, pp. 996-1011, 1996.
[437] S. Baker and T. Kanade, "*Hallucinating Faces*," 4th IEEE International Conference on Automatic Face and Gesture Recognition, pp. 83-88, France, 2000.

CHAPTER 8

CONCLUSION AND FUTURE WORKS

8 Chapter 8: Conclusion and Future Works

8.1 Conclusions

Having a video sequence as the input, this Ph.D. thesis goes into the details of the different steps of a system that aims to produce a high-resolution and high-quality frontal face image from the input. These steps are face detection, face quality assessment and face quality improvement.

For the face detection, it has been tried to develop a new face detector based on the color information and the facial features. This face detection is slower than the state-of-the-art face detectors. However, it can detect the profile face images and produces better detection rates.

In the face quality assessment part different face quality measures, their normalization and their conversion into quality scores are explored using different methods. The results of the face quality assessment have been used to find best face-logs, and complete face-logs. Then, some similarity measures are employed to evolve the complete face-logs into over-complete ones. These over-complete face-logs are then passed over to the enhancement part of the system.

In the enhancement part of the system, super-resolution algorithms are used to produce a high quality face image from the over-complete face-logs. Meanwhile, different types of super-resolution algorithms and their pro and cons are explored. Then, a new hybrid super-resolution algorithm for face images has been introduced and it is shown that this algorithm can operate very well on the over-complete face-logs. Furthermore, a face recognition algorithm has been developed for working with long video sequences and it is discussed that if the input of this recognition algorithm is obtained by our hybrid super-resolution algorithm from the over-complete face-logs, the recognition rate of the system can be improved.

Several different local and public databases have been used for testing different parts of this system throughout this thesis. Changes in the values of the different involved facial features in these databases ensure the reliability and efficiency of the proposed system.

8.2 Future Works

The proposed work in this thesis has been applied to the video sequences containing face images however, the overall idea of the quality assessment, finding key frames and improving their quality can be applied to different computer vision applications.

In the face detection part, combining the fast Viola & Jones face detector with some information of the image like human skins can be promising. Furthermore, its false positives can be reduced by employing some facial features like dimension of face images etc. Obviously, there should be always a compromise between the speed of the face detector and involvement of such heuristics.

In the facial feature extraction part, more studies can be conducted to find facial features that are more robust. Especially, reliable and invariant features that can be extracted in low-resolution images are desirable.

Pursing face-log generation for other purposes (than video indexing, recognition, and super-resolution) can lead to new methods for combining the score and consequently new face-logs.

The hybrid super-resolution algorithms introduced in chapter 6 and 7 seems very promising. Trying different combinations of reconstruction-based and recognition-based super-resolution algorithms can produce very interesting results for many computer vision applications.

COMPUTER VISION & MEDIA TECHNOLOGY LABORATORY

The Computer Vision and Media Technology Laboratory is part of the department of Architecture, Design and Media Technology. CVMT is founded in 1984 as Laboratory of Image Analysis (LIA) at Aalborg University.

The main research areas of the laboratory are:

- Computer Vision
- Multimedia Interaction
- Virtual Reality
- Augmented Reality
- Autonomous Systems and Agents

CVMT has established research cooperation with more than 30 universities in 16 different countries.

CVMT is currently headed by Associate Prof. Hans Jørgen Andersen.

CVMT

COMPUTER VISION AND MEDIA TECHNOLOGY LABORATORY

AALBORG UNIVERSITY

NIELS JERNES VEJ 14

DK-9220 AALBORG

DENMARK

TELEPHONE: +45 9940 8793 TELEFAX: +45 9940 9788

URL: HTTP://WWW.CVMT.DK

ISBN: 978-87-992732-5-6